AD-A100 470 TECHNOLOGY SERVICE CORP SANTA MONICA CA F/G 5/8
MULTISTATION VOICE DATA ENTRY CONFIGURATION STUDY.(U)
APR 81 P W GREGORY, J M REAVES F30602-80-C-0033
UNCLASSIFIED TSC-PD-B662-1 RADC-TR-81-50 NL

1 of 2
AD
A100470

(12)

AD A100470

RADC-TR-81-50
Final Technical Report
April 1981

MULTISTATION VOICE DATA ENTRY CONFIGURATION STUDY

Technology Service Corporation

Peter W. Gregory
J. Michael Reaves

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-81-50 has been reviewed and is approved for publication.

APPROVED: Richard S. Vonusa

RICHARD S. VONUSA
Project Engineer

APPROVED: John N. Entzminger

JOHN N. ENTZMINGER
Technical Director
Intelligence and Reconnaissance Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (IRAA) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER RADC-TR-81-50	2. GOVT ACCESSION NO. AD-A100 470	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MULTISTATION VOICE DATA ENTRY CONFIGURATION STUDY		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report Jan 80 - Oct 80
		6. PERFORMING ORG. REPORT NUMBER TSC-PD-B662-1
7. AUTHOR(s) Peter W. Gregory J. Michael Reaves		8. CONTRACT OR GRANT NUMBER(s) F30602-80-C-0033
9. PERFORMING ORGANIZATION NAME AND ADDRESS Technology Service Corporation 2950 Thirty-First Street Santa Monica, Los Angeles County, CA 90405		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 64701B 43030341
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (IRAA) Griffiss AFB NY 13441		12. REPORT DATE April 1981
		13. NUMBER OF PAGES 142
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Same		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Richard S. Vonusa (IRAA)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Voice Data Entry (VDE)	Speech Synthesis
Automatic Speech Recognition (ASR)	Voice Response
Voice Input/Output	On-Line Data Entry
Voice Recognition	Cartographic Analysis
	Distributed Computer Systems

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes a study to configure a multiple-station voice data entry (VDE) system for application to the Digital Landmass System (DLMS) data base at the Defense Mapping Agency's Aerospace Center (DMAAC) in St. Louis, Missouri. Study results are based on a thorough analysis of the DLMS data entry process and the analysts' environment, including a detailed questionnaire completed by over 80 analysts. (A tabulation of questionnaire responses is given in Appendix B.) Available equipment for

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) voice data entry and response was surveyed extensively; results ~~were tabu~~lated and are presented ~~in Section~~ 3. An operational scenario for VDE is developed ~~in Section 6,~~ following a detailed functional description of DLMS data entry requirements ~~(Section 5)~~. Finally, detailed multi-station configurations are presented in Section 7. Both highly centralized and fully distributed configurations meet system requirements; for several reasons, however, we consider recommendation of specific hardware inappropriate.

The advantages of VDE system are complicated by considerations of cost-effectiveness and the current procedures of data entry. The small percentage of analyst time actually spent on data entry makes cost a major issue. Furthermore, it was difficult to estimate the advantages of interactive, on-line data entry by voice over interactive, on-line data entry by keyboard, because present data entry is performed entirely off-line.

Interactive, on-line data entry by keyboard is due to be implemented at DMAAC with IFASS (Interactive Feature Analysis Support System). This and other advanced systems being actively sought by DMA will substantially change the analysts' data entry procedures. We therefore recommend that VDE be reevaluated with respect to specific hardware cost/performance trade-offs after IFASS is implemented and in light of the rapid advancements being made in speech recognition and voice response technology.

EXECUTIVE SUMMARY

STATEMENT OF THE PROBLEM

Creating the Digital Landmass System (DLMS) data base at the Defense Mapping Agency Aerospace Center (DMAAC) involves a great deal of hands-busy, eyes-busy effort. The analyst's primary task, photo interpretation, requires that data be entered into a computer in a specified numeric format--a process that interferes with the primary task and breaks the analyst's concentration. A voice data entry (VDE) system should simplify the data entry process, leaving the analyst's hands and eyes free to concentrate on interpreting photographs.

An advanced development model (ADM) of a particular VDE system has been installed at DMAAC and has undergone preliminary evaluation tests. These evaluation tests yielded mixed results, pointing up that properly implementing and integrating VDE devices into a total system is essential to their eventual success. And for these devices to be most cost-effective, an optimum configuration and sharing of resources must be developed.

The ADM provided DMAAC with their first real experience of voice data entry. For this reason, we have carefully examined DMAAC's evaluation tests on it. We have also sought out analyst reaction to the concepts and procedures of VDE in general, as distinct from their implementation in the ADM. Finally, we have attempted to remain objective and unbiased toward the particular advantages and disadvantages of a multiple-station voice data entry system for DMAAC. This report examines the DLMS data entry problem in terms of a multiple-station (multistation) VDE configuration.

NATURE OF THE DATA ENTRY PROCESS

At present, over 100 analysts are participating in the creation of the DLMS data base at DMAAC by entering data and analyzing features. The feature-analysis task is performed using light tables and a stereoscopic viewer on several photographic and cartographic sources to create overlays and identify feature characteristics. Data entry, an off-line process, consists of completing a Feature Analysis Data Table (FADT) by determining appropriate numeric codes for the various feature parameters. This numeric data table is either entered onto special forms for optical scanning or punched onto IBM cards, which then get entered onto the main Univac 1108 computer for compilation, verification, and storage.

Both methods of data entry have their own advantages and disadvantages. Feature data can be written onto optical-scanner (OPSCAN) forms directly at the analyst's workstation, and information on as many as ten features (FACs) can be described on a single sheet. However, the optical scanner requires that the numerals be drawn on a special six-segment matrix in an unnatural, highly structured manner using a #1 graphite pencil. All too frequently the optical scanner will be inoperable, and even when it is working properly, the nature of the forms, the restricted manner of writing numerals, and smearing tend to cause errors.

Keypunched data entry is preferred by some analysts because it is simpler and generally results in fewer errors. The analyst can specify features at the workstation, with some shorthand notation and then take this data to be keypunched. Keypunching personnel are available at DMAAC but are not often used in practice because of the shorthand nature of the analyst's data (i.e., the analyst would have to spell everything out explicitly on keypunch forms for the keypunch operators) and the fact that the analysts can spot mistakes by reviewing their own work as they keypunch it. A shortcoming of this practice is that there are only a few keypunch machines available, and they are not always working perfectly. Although several analysts insist that keypunching is a faster data entry mode than OPSCAN entry, analysis skills are obviously being wasted by time taken to keypunch data.

Each analyst has his/her own methods of identifying and entering data, and these methods vary with time and task. Generally, most analysts will perform a significant amount of analysis (on the order of a half hour or more) before entering any data, but the exact amount depends, to a large extent, on the particular area being analyzed. Of course, since the system is off-line, these data are not really "entered" anywhere, but merely specified on FADT forms for scanning or keypunching at a later time, often after the entire area or manuscript has been analyzed. An entire manuscript might take several weeks or several months to complete and could consist of as many as two or three thousand features, corresponding to a stack of FADT forms several inches thick.

Verification of data is not performed in real time. It consists of indicating errors (illegal values for parameter codes) or missing or extraneous values and indicating unusual or suspect features (combinations of parameter codes not normally encountered). Verification is performed feature by feature; so, correspondence and overlap between features are not verified. After the data have been entered, verified, and stored on the computer, the FADT forms will generally be saved for a year or more in case the data get lost or damaged on the computer.

ASPECTS OF THE PROBLEM

As a result of Technology Service Corporation's (TSC's) visits to DMAAC and the TSC questionnaire completed by 84 of the analysts, several key aspects of the data entry process were identified:

1. Present procedures are wholly geared to off-line, non-real-time analysis, data entry, and verification. The analysts have dealt with this situation by developing their own individual techniques for analysis and data entry, to make the best use of their time and skills. Any on-line data entry device, regardless of whether it makes use of voice, presents a fundamental change to the present system and will upset many of the analysts' present procedures. The introduction of on-line data entry and verification will, therefore, require an adequate transition period so that new procedures can be established to take full advantage of this new capability.
2. Data entry requires only a small percentage of an analyst's time, and since it is performed at a relatively slow rate, entry speed is not now a critical factor. Most DMAAC personnel agree that data entry requires only 10 to 15 percent of an analyst's time. Data entry rates vary from one entry every 2.5 minutes to just over two entries a minute, on the average, during peak periods. These average data entry rates have little meaning, however, considering that data are entered off-line and that analyst data-recording procedures vary. In fact, from a small sample of data taken during the ADM evaluation tests, it was concluded that simulated analysis and data entry using OPSCAN sheets versus on-line voice or keyboard entry resulted in similar performance times--again, not surprising in view of the present analysis procedures. Considering these points, one could conclude that the major justification for on-line data entry would be improved accuracy of the resulting digitized data; however, no formal tests of the potential for improvement in accuracy of the overall process have been attempted.
3. Given the analyst's environment, crowded and cluttered with various analysis paraphernalia, the advantages of hands-free, eyes-free data entry should be great. But, evaluation tests of the ADM VDE system failed to confirm this advantage. In fact, several disadvantages of the present ADM pose serious drawbacks to implementing such a VDE system: 1) the requirement that a headset be worn, and the headset's discomfort and inconvenience to the analyst; 2) the lack of real-time hardcopy output, resulting in analyst confusion and inability to either check similar features or compare related ones entered previously; 3) the unnatural, halting manner of speaking required for consistent recognition accuracy; and 4) the difficulty of the system to meet security requirements (i.e., TEMPEST).

Thus, examination of the data entry problem resulted in the conclusion that voice data entry has not yet been demonstrated to have significant advantages over other on-line, real-time data-entry and -verification systems, and, in any event, the present ADM VDE system does not meet DMAAC needs.

CONFIGURATION APPROACHES

Early in the project, TSC was directed not to limit configuration analysis to equipment incorporated into the ADM. From discussions with personnel at DMAAC and Rome Air Development Center (RADC), TSC concluded that a broad range of interactive multistation configurations should be examined, including those making only limited use of commercial voice recognition and response capabilities. We refer to these as mixed-mode configurations because data entry can be performed either by voice, or manually, or by a combination of the two. Mixed-mode configurations can take maximum advantage of the different capabilities of keyboard- and voice-entry modes, and free the analyst from the restrictions that either entry mode alone would impose. Thus the analyst is allowed the freedom to develop procedures and techniques that will work best for the varying analysis tasks encountered.

In performing the configuration analysis, we took into account the specification and procurement of two advanced systems for DLMS data entry and analysis already underway at DMA. The first system, IFASS (Interactive Feature Analysis Support System), is an on-line, keyboard-type data entry terminal to be used by each analyst. The second system, CAPI (Computer Assisted Photo Interpretation), will greatly simplify the analyst's photo-interpretation tasks. Information on these systems is limited, but we have attempted to make our study flexible and general enough to be compatible with them.

After generating data flow diagrams for the DLMS data base and VDE processes, we examined three basic configurations. Each configuration can be defined by partitioning the appropriate data flow diagram at different points.

The first configuration is highly centralized, with a central minicomputer performing all data entry and processing and storage tasks. Resources are highly shared, making this potentially the most cost-effective configuration. However, the cost of minicomputer-based full-scale implementation of VDE is great, and becomes the overriding cost factor. Also, this configuration has the least flexibility for expansion, the greatest reliance on the central unit, and hence the greatest sensitivity to faults and failures of the central unit--and potentially the worst response time for voiced entries.

The second configuration we examined is partially centralized on two levels; that is, three central computers perform data entry processing and storage functions. Word reference patterns are stored in several subsets, and voiced entries are recognized at each user's station. The data flow between the central computer, the two station controllers, and user stations now becomes less critical, response times are better, and limitations on reliability and flexibility are not as severe. This configuration requires intelligence at each station for the speech recognition processes, and it also allows for some local backup memory to reduce its dependence on the central computer.

The third configuration consisted of independent stations for each analyst. Analyst-to-analyst interaction is not required, so there is no fundamental reason to centralize anything, which means that each analyst will have his own program and storage capabilities. This configuration requires redundant data entry programs at each station, but it also yields the highest reliability and flexibility: Since each station is independent, additional stations can be added at will; and since the processing and storage requirements of a central computer have essentially been distributed among the independent stations, there is no central processor to fail and halt the operation of all user stations. When a user station fails, it will affect only that particular station, and this, of course, is advantageous in a production environment. However, the cost-effectiveness of an independent-station configuration will depend on whether low-cost components for each station can be found.

These basic configurations are described and illustrated in Section 7. Hardware costs for all three configurations are high compared with just keyboard entry alone, and although the independent stations configuration shows the greatest promise for mixed-mode voice data entry, all three configurations are, in fact, viable.

CONCLUSIONS

This configuration study presented both unique problems and unique opportunities in examining a potential application of voice data entry (VDE). The usefulness of data entry and verification by voice would seem to be obvious in the labor-intensive, manual, off-line processes now used for data entry to the DLMS data base. Completing OPSCAN forms and keypunching are poor examples of man-machine interaction and wasteful of the analyst's unique skills. That the present system of data entry is outmoded and in need of improvement is evidenced by the Interactive Feature Analysis Support System (IFASS) and the Computer Assisted Photo Interpretation (CAPI) system due to be implemented at DMAAC.

Our conclusion, however, is that the use of voice data entry is not an all-or-nothing choice and that full-scale implementation of VDE is unwarranted at DMAAC at this time. This conclusion is based on the results of this configuration study, as well as the experience of DMAAC with the advanced development mode (ADM) system.

Judging from the ADM evaluation report, it would appear that keyboard data entry is superior to VDE and preferred over it. But the evaluation tests consisted of constrained tasks and a small sample size, and we conclude that further tests of the effectiveness of VDE need to be performed. We further conclude that such additional testing can best be accomplished using the on-line data-entry and -analysis capabilities of IFASS and CAPI.

The technological leap to VDE from off-line, manual data entry has been successfully made in a number of production environments. But dependence on VDE is neither appropriate nor cost-effective in this DMAAC application. Currently, data entry comprises only a small portion of the analyst's time, and VDE equipment poses practical problems in the DMAAC environment, as discussed above. On-line keyboard data entry offers significant improvements relatively inexpensively. It presents the best alternative at this stage and is recommended.

Any advantages of VDE over keyboard data entry will become apparent only after on-line data-entry and -verification procedures have been established, and after further experience is gained with VDE capabilities. We believe that both these goals can be met with a mixed-mode data entry system that will allow keyboard data entry as well as limited-capability VDE. Although the advantages of full-scale VDE implementation will not be available, the cost and training requirements of mixed-mode VDE will be low enough to allow for implementation, test, and experience on a much larger scale. Without such experience, the practicality of VDE for DLMS data base entry is unclear and subject to conjecture.

RECOMMENDATIONS FOR FUTURE WORK

Our recommendations for future work are as follows:

1. Develop procedures for on-line, real-time data entry and verification for the DLMS data base, including the specification of probable data entry rates and realistic data response times. The development of such on-line data entry procedures will require a thorough examination of analysis tasks to arrive at the best analyst-computer interaction.

2. In conjunction with the IFASS keyboard/keypad equipment, design a mixed-mode data entry station using state-of-the-art speech recognition technology. This station will be able to accept voice and keyboard data interchangeably, at any time, according to analyst desires.

3. Implement and test limited-capability (20-to-30-word vocabulary) VDE for data entry speed and data accuracy in DFAD compilation. Feature ID codes will be entered as codes rather than as verbal descriptions, thus improving recognition accuracy while greatly reducing analyst training and retraining times.

4. Evaluate comfortable, nonobstructing earpiece microphones as replacements for the close-talking headset-mounted microphones typically required for VDE equipment. If they perform acceptably, such hearing-aid-type microphones could provide a considerable advantage for VDE in Defense Mapping Agency applications.

5. Examine high-speed, high-quality voice response technology for prompt, feedback, confirmation, and verification functions associated with VDE. If designed and implemented properly, voice response should improve data entry speed while reducing the number of errors.

CONTENTS

CONTENTS (Cont'd)

LIST OF FIGURES

LIST OF TABLES

1. BACKGROUND

The Defense Mapping Agency Aerospace Center (DMAAC) is responsible for meeting the aerospace mapping, charting, and geodesy (MC&G) requirements of military organizations in the Department of Defense. DMAAC is oriented toward production, particularly production of geographic data files. One of its principal products is the Digital Landmass System (DLMS), which consists of two data files, the Digital Terrain Elevation Data (DTED) file and the Digital Feature Analysis Data (DFAD) file.

DLMS is used by both internal organizations and external users in such applications as visual simulation, radar simulation and production, electro-optical visual system (EVS) simulations, automated cartography, mission planning, and radiometrics. The DTED file contains terrain elevation data in matrix format. The DFAD file contains digitized information on cultural features. Unlike the DTED, for which data are collected at standard intervals, the DFAD has data collected in varying densities according to the cartographic features of specific areas. Cultural data are organized within culture manuscripts, encompassing specific geographic regions.

In this section, the creation of the DLMS data base and the data entry environment at DMAAC are discussed briefly. The theoretical advantages of voice data entry (VDE), as well as the practical disadvantages of present VDE equipment, are introduced. The experience of DMAAC with an advanced development model (ADM) VDE system is described at some length. Lastly, plans for improved analysis and data entry systems at DMAAC are introduced.

1.1 DLMS DATA BASE

DLMS culture manuscripts are created as follows. Planimetric data are obtained for an area from one or more sources, e.g., aerial photography, etc. They are then oriented, analyzed, and combined to form the geographic scenario that is stored in DFAD. Each feature has several pieces of information about it encoded and stored, including a Feature Analysis Code (FAC); the feature's type; code identifying what feature it is; its predominant height and surface material; and, depending on its type, other information such as length, width, orientiation, structure and tree densities. A description in detail of DLMS feature attributes appears in Appendix A.

The descriptive information for each feature is encoded in the Feature Analysis Data Table (FADT) according to the Product Specifications for the Digital Landmass System Data Base [Defense Mapping Agency 1977].* The FADTs are coded forms from which the data are keypunched or optically scanned and then input into the DMAAC batch computer (Univac 1108). The physical-feature outlines are digitized from a geographical plot in the order of their FAC numbers and combined in the computer to form the DFAD file. The data are then subjected to verification, error correction, and merging with other manuscripts, and eventually become a new DLMS tape.

1.2 DMAAC DATA ENTRY ENVIRONMENT

Over a dozen analyst sections work on DLMS, each section consisting of about ten analysts. Each section has its own area, and some sections are located in adjoining rooms. Each room is filled to capacity, and the placement and storage of any additional equipment requiring floor space would be difficult.

Each analyst sits in a swivel chair and is surrounded on three sides by a desk, a light table, and a stereoscopic viewer mounted on a light table. The stereoscope table is portable and can be moved to other areas as needed. Analysts use a variety of equipment at their workstations, including seven different colored pencils, a sharpener, acetate, paper, FADT forms, a graphite pencil, a magnifier, a calculator, a DLMS specification for code lookup, and some one-page summary sheets of the more popular codes.

Analysis is a visually demanding task, requiring a high degree of concentration. Physically, the analyst must manipulate the stereoscopic views by means of a stop-action clutch while manually positioning and holding large sheets or rolls of photographic and other source material. Any physical motion, for example, swiveling around to the desk to check on a feature identification code, will tend to knock the source material out of alignment, and correct positioning must be regained.

Analysts will occasionally leave their workstations to confer with other analysts working on similar areas. In general, the environment is noisy with the sounds of people talking, papers being shuffled, chairs squeaking, and miscellaneous sounds.

---

*Defense Mapping Agency, Product Specifications for Digital Landmass System (DLMS) Data Base, Defense Mapping Agency Aerospace Center, St. Louis AFS, Missouri, July 1977.

1.3 ADVANTAGES/DISADVANTAGES OF VOICE DATA ENTRY

Data entry by voice offers a large number of potential advantages over data entry by conventional means. Speech is man's most natural mode of communication and is performed easily by nearly everyone. Entering data by voice into a machine allows an operator's hands and eyes the freedom to perform other tasks at the same time. Data can be more quickly spoken than written or typed, and speaking can be performed without laborious training such as that required to attain proficiency in typing.

The man-machine interface is shifted by voice data entry (VDE) to accommodate the human rather than the machine. For example, real-world items typically have to be described for a computer in terms of numbers, but a VDE system can accept verbal descriptions and translate these internally into the proper numeric codes. A voiced entry provides inherent feedback to the operator because it is spoken, and entries can be monitored simultaneously by a third party, if desired. In the DLMS application, this monitoring capability presents a distinct disadvantage in light of security requirements. Computer entries can be made in a single step by each user, or, alternatively, after a complete series of entries have been recognized and verified. Data entry by voice can provide more accurate entries and hence less rework time from errors.

Not all of these advantages are important or even pertinent to every application. And the possible advantages of speech recognition must be traded off against possible disadvantages of present-day VDE equipment.

Most VDE systems today require each user to train the system for his/her voice for every word in the vocabulary, and to store these reference patterns for future use. Depending on the size of the vocabulary, this process can take hours and still require retraining later because of subtle voice changes or the physiological effects of a cold or hay fever. Thus, each user must become adept at speaking in a consistent manner such that his/her day-to-day utterances match the trained reference patterns.

Most VDE devices require isolated words or phrases to be spoken as entries. That is, an artificial pause must separate each distinct entry, and each pause must be significantly longer than the natural pauses that occur within words or between words in a phrase. For example, "822" would have to be entered as "EIGHT" pause, "TWO" pause, "TWO" pause. This pausing leads to an unnatural, halting mode of speech, which is difficult to master. Thus, both the number of speakers and their manner of speaking are restricted by present-day VDE systems. The few devices that provide exceptions to these restrictions are described in Section 3.1.

In addition, VDE devices require a vocabulary of limited size and complexity. That is, if the machine finds that someone's pronunciation of two words is too similar for accurate recognition, one of the words will have to be changed. Such machine confusions are not necessarily obvious to human listeners; for example, "FIVE" and "NINE" are traditionally confusing words for automatic speech recognition and sometimes result in recognition errors. By the same token, recognition accuracy, even under the best conditions, is not perfect, although a threshold can usually be set so that the machine can reject a word it is unsure of rather than take the chance of misrecognizing it. Machine rejection does not solve the problem of limited recognition accuracy, however. Generally, the larger the vocabulary, the lower the recognition accuracy and the longer the response time between an utterance and the machine's response to it.

To obtain maximum recognition accuracy, most VDE devices require that a close-talking, noise-cancelling, headset-mounted microphone be used. But such headsets are rarely comfortable when worn over long periods of time, and the cord and the boom-mounted microphone can become snagged or knocked out of position with operator movements.

Another significant factor for user acceptance is the psychological aspect of using a VDE system. While many users react favorably to this new technology, others fear that it will take away part of their responsibility, or see it as automation merely for the sake of automation. Some people think it is degrading to have to talk to a machine, and others object to having to hear themselves (and other users in the area) talk all day long. Thus, in addition to limitations on the VDE equipment itself, the concept of data entry by voice is often viewed in a negative light by users, especially when users are accustomed to performing data entry another way and do not see any obvious advantages to a VDE system. User acceptance is discussed further in Appendix B.

1.4 THE ADVANCED DEVELOPMENT MODEL SYSTEM

A major step toward examining VDE for the DLMS data base was taken with the implementation of a system at DMAAC incorporating Threshold Technology's

VIP-100 equipment. This advanced development model (ADM) system is described fully in the report by Scott [1980].* The basic configuration, Figure 1, consists of a full-sized rack of equipment and several peripheral devices, including a keyboard CRT terminal and a visual response unit integrated into the Bausch & Lomb Zoom 240 stereoscopic viewers.

The VOTRAX ML-1 voice response unit did not seem to be working properly, either during an informal demonstration of the system in March 1980 or afterwards: There were long and unpredictable pauses between voice input and the unit's response; and the quality of the VOTRAX voice outputs was not very good even under the best of conditions.

Voice inputs were accepted through a Shure Brothers SM-11 headset-mounted microphone. This is an exceptionally high-quality, light-weight device, but a nuisance to wear over long periods of time and while performing other tasks. To avoid the discomfort of having to wear a headset, a fixture was added to the stereoscopic viewer that enabled the microphone to be mounted on it. This alternative turned out to be just as inconvenient, however, because it required the analysts to always be right at the stereo viewers to enter data.

ADM performance was marred by several hardware problems, including electrical interference from the light tables and problems with the XEBEC floppy disk drives. Nonetheless, evaluation tests were performed on the ADM, presumably with all the equipment working properly.

The evaluation tests were performed in two parts. The first part consisted of two experienced (CDI) analysts entering data by OPSCAN sheets, keyboard, or voice after they had derived FADT information on 50 preselected features, assigning preliminary Feature Analysis Codes (FACs), compiling an intermediate working overlay from rectified photography, and assigning unique (final) FAC numbers. The second part consisted of seven other (CDV) analysts copying the FADT data derived previously by the CDI analysts, using the same three methods for data entry: OPSCAN sheets, keyboard, and voice.

Only total task times were recorded for each analyst. Time required for training and retraining of the 248 words that made up the VDE vocabulary was not counted, as it would have extended voiced data entry times excessively and, it was agreed, unfairly. Information was not available on data entry rate and

---

*Phillips B. Scott, Final Technical Report: DLMS Voice Data Entry, Contract F30602-78-C-0327, Rome Air Development Center, Griffiss AFB, New York, 1980.

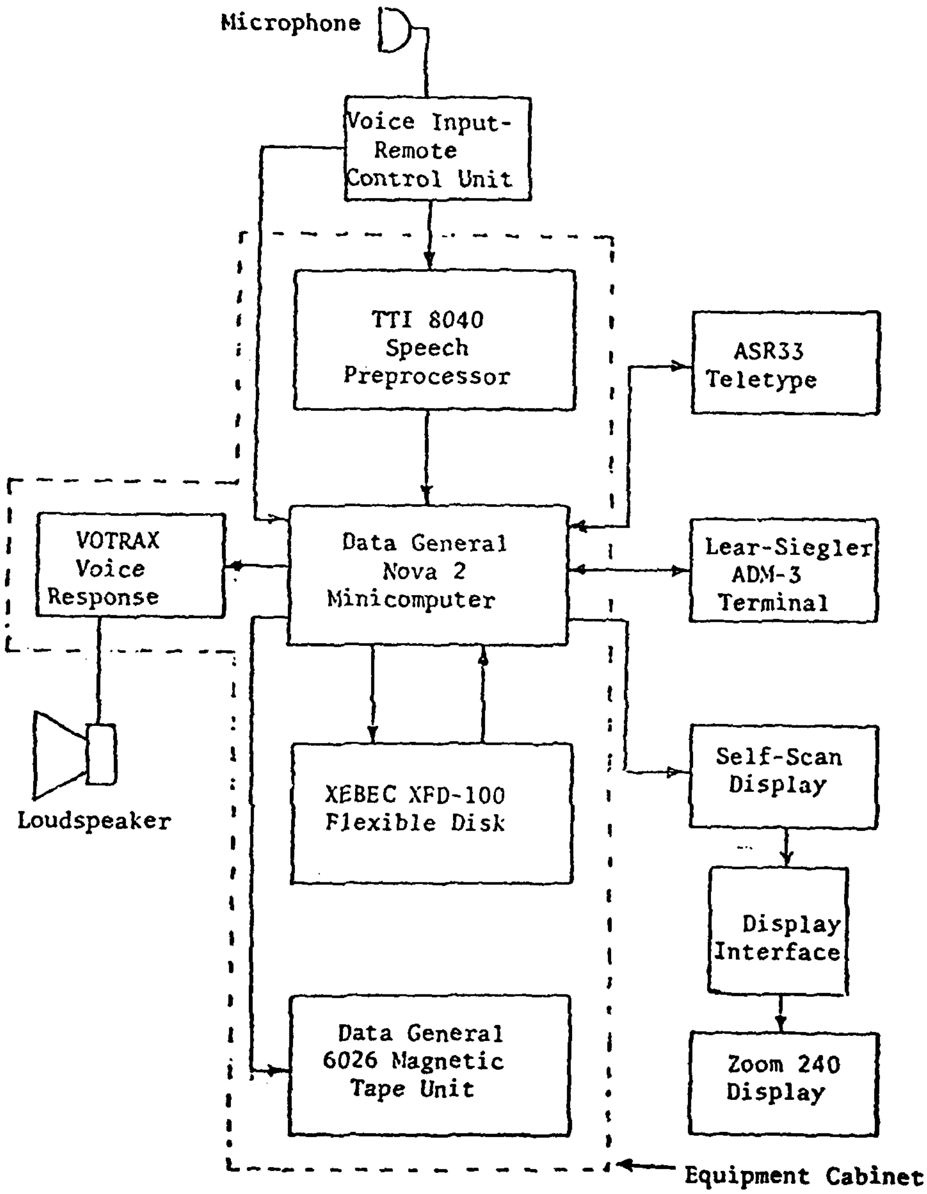


Figure 1. Block diagram of DLMS voice recognition system.

variability; that is, the amount of time spent on analysis and how it was punctuated by data entries was not kept track of, but only average entry rates. On the average, three entries were made every minute of compilation time, regardless of the method of data entry. For data copying, OPSCAN performance averaged 7 entries per minute, voice averaged just over 12 entries per minute, and keyboard averaged 20 entries per minute.

It is important to understand that the compilation and copying tasks were quite different, and conclusions based on one cannot be applied directly to the other. The two tasks are compared in Figure 2 by average total performance time per entry mode. The task of entering a mass of data is very different from the task of determining what data are to be entered and then entering them bit by bit, as they are determined.

From the results of the data copying task, it was concluded that keyboard entry is 38 percent faster than voice, which is not surprising. In fact, this result agrees with that found by Threshold Technology (TTI) three years ago during some comprehensive tests performed for RADC [Welch 1977].* In what were described as high-speed data entry tests, TTI found that keyboard entry was 29 percent faster than voice just for copying data, and resulted in fewer errors when the data were strictly numeric.

That result was reversed (voice entry was almost 30 percent faster than keyboard) when the simple data copying task was replaced with a complex data entry task that included other ongoing tasks. Hence, the speed of VDE versus that of keyboard entry was dependent on the particular task involved. So, although keyboard entry was faster than VDE, and voice was faster than OPSCAN for the data copying (CDV) task, neither entry mode can be considered more advantageous in the more realistic compilation (CDI) task.

Results for the CDV and CDI tasks are compared in Figure 3. Individual performance times are shown in relation to the average OPSCAN time for each task (see Figure 2). That is, the average OPSCAN performance times (68.6 minutes for the CDV task including rework, and 184 minutes for the CDI task including rework) became the standard and were set to 100 percent in the figure. All other performance times, including average keyboard- and voice-entry times, are expressed as a percentage of the OPSCAN time.

---

*J. R. Welch, Automatic Data Entry Analysis, Final Technical Report RADC-TR-77-306, Rome Air Development Center, Griffiss AFB, New York, 1977.

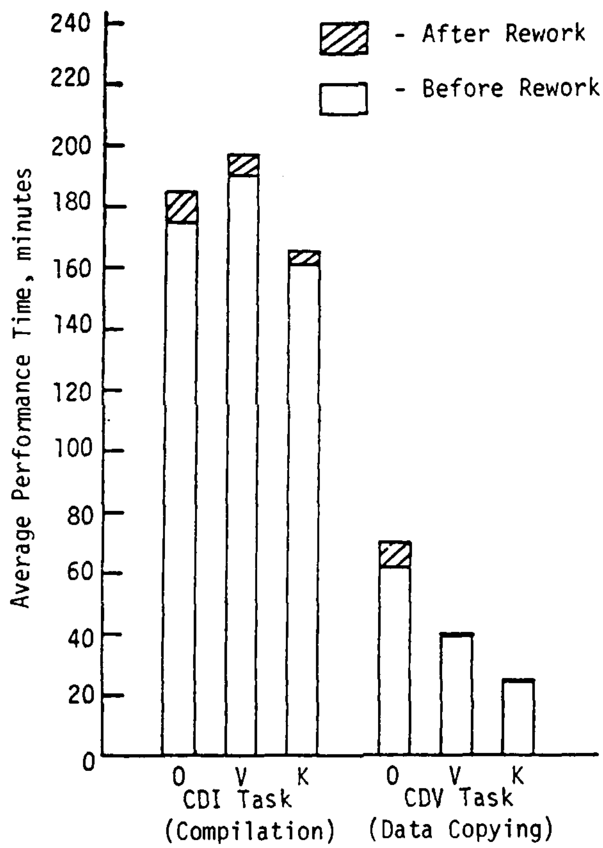


Figure 2. Comparison of CDI and CDV tasks by average performance time for three modes of data entry (O= OPSCAN, V = voice, K = keyboard).

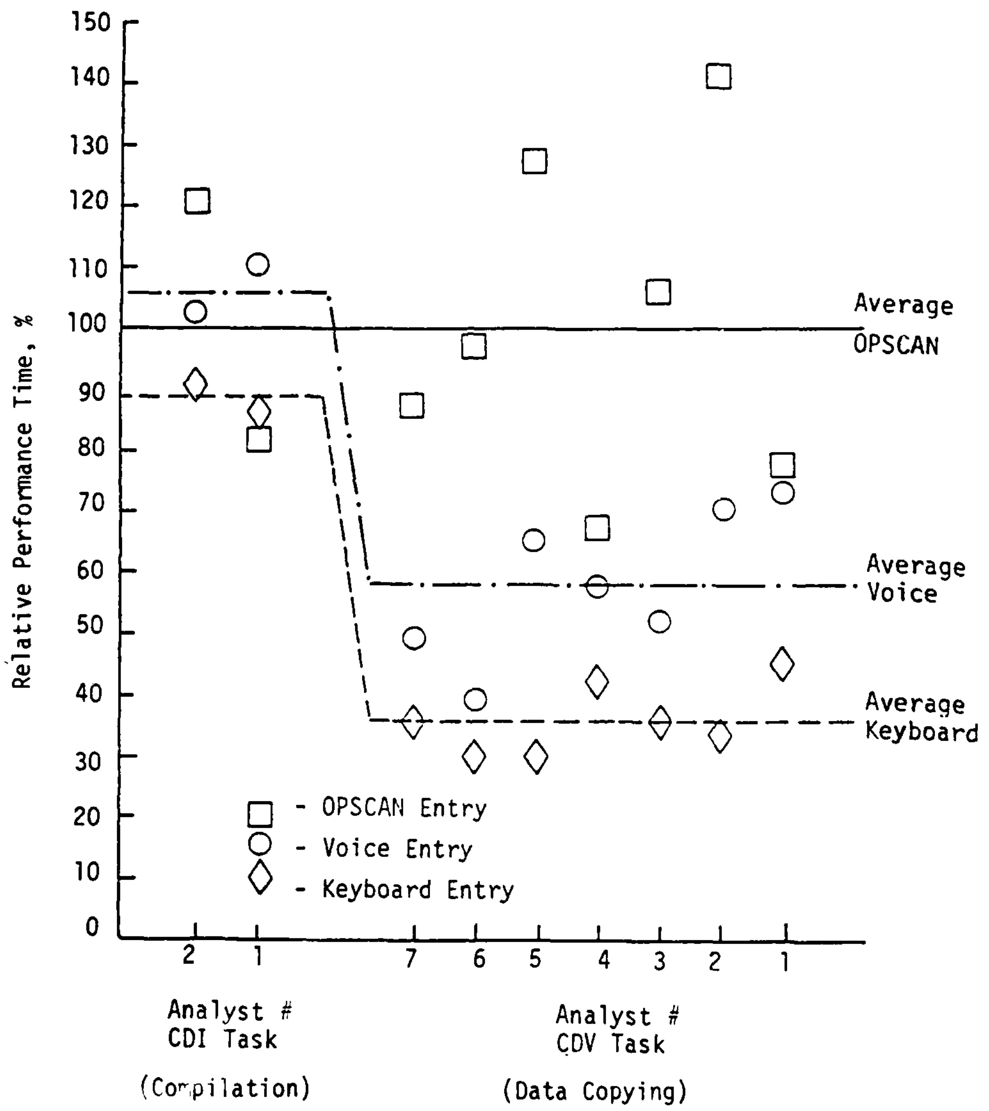


Figure 3. Comparison of individual performance times relative to the average OPSCAN times for CDI and CDV entry tasks.

Not only do the average performance times for voice and keyboard entry vary greatly relative to the average OPSCAN time for the two tasks, but the analyst-to-analyst variation is also large. OPSCAN was the slowest mode of data entry for one of the two CDI analysts performing compilation, but the fastest for the other. The data in Figure 3 also indicate that the analyst-to-analyst variation was greater for OPSCAN than for voice or keyboard data entry. In other words, for both the CDI and CDV tasks, voice and keyboard entry resulted in more consistent performance times.

Neither keyboard nor voice data entry can be claimed to be significantly faster than OPSCAN, however, owing to the small sample size for the CDI task. Until further testing is done using voice, keyboard, and OPSCAN for a realistic analysis task, it remains to be seen whether keyboard or voice has a significant advantage in speed over OPSCAN data entry. Regardless, speed should not be the only criterion used for comparison. Other factors such as ease of use and operator fatigue should also be considered, although these factors are difficult to measure quantitatively. Additional experience with the ADM should provide such qualitative feedback.

1.5 IMPROVED SYSTEMS AT DMAAC

Two improved systems are scheduled for implementation at DMAAC. The first is an on-line, interactive keyboard terminal and display station system referred to as IFASS, Interactive Feature Analysis Support System. The second is a feature analysis/digitizing system called CAPI, Computer Assisted Photo Interpretation.

IFASS will consist of interactive CRT keyboard terminals with redundant keypads for data entry at the workstation. The terminal will connect to a central computer, probably a minicomputer. Each terminal will be used for communication with the central computer, but will have no separate processing capability of its own; that is, it will probably not include any programming intelligence.

The CAPI stations will drastically reduce the analyst's workload by automatically assigning FAC numbers and allowing descriptive data to be associated with a feature as it is digitized. Just as the capabilities of CAPI are much greater than those of IFASS, its cost is also likely to be much greater. Tentative plans are to procure four IFASS systems, each having 42 analyst stations, and 70 CAPI workstations.

The IFASS system is of greatest interest to this configuration study because it provides for on-line, real-time data entry and verification. With proper software, IFASS could provide a wealth of information on analysis compilation, data-entry, throughput, and error rates. Such information is just not available today with the present off-line data-entry and verification procedures, and the lack of such information presents a serious impediment to configuration analysis. As discussed in Section 2, average data entry rates have been estimated from information provided by DMAAC, but this information is no substitute for actual data-entry and error rates which could be compiled as a by-product of IFASS.

IFASS could also provide a multiple-station structure for testing voice data entry capabilities. More than one VDE manufacturer makes speech recognition equipment that fits inside a standard CRT keyboard terminal and is essentially invisible to the user. Most VDE manufacturers offer equipment that is RS-232C compatible, making it possible for their speech recognition device to be plugged into each IFASS terminal, assuming these terminals have such an interface to spare. In any event, the IFASS specification requires that the CRT terminals have available an extra logic-board slot for the addition of a secondary input device, such as that provided by VDE capability. Various data entry structures, operating procedures, and verification, error correction and data storage capabilities could be evaluated without the need to purchase expensive, stand-alone, full-capability voice data entry hardware.

Thus, IFASS presents several unique opportunities for evaluating voice data entry for the DLMS data base, and these will be presented further in the recommendation section of this report (Section 8).

2. DATA FLOW REQUIREMENTS

Understanding how data moves in the DLMS process is essential to this configuration study. Data flow diagrams will thus be used to guide the design of the multistation configuration as well as to define specific requirements of the configuration for internal functions and interfaces to the outside world. These diagrams were generated from discussions with DMAAC personnel, visits to the DMAAC facility in St. Louis, Missouri, and reviews of various documents describing DLMS processes.

2.1 DATA FLOW IN THE PRESENT SYSTEM

The data flow diagram of how the DLMS data base is created, Figure 4, shows that various image and textual sources, including maps and photographs, are used to generate culture (DFAD) and terrain (DTED) tapes. These tapes, combined, form the DLMS data base. DFAD generation comprises feature extraction and verification, and the digitization of feature coordinates. The generation of FADT data is shown in detail in Figure 5 as part of the extraction of feature data, which was represented as a single node in Figure 4. Note that specific processes or tasks are not shown, only the steps through which data pass.

FADT generation proceeds as follows. Features are selected from the manuscript area, and temporary FAC (Feature Analysis Code) numbers are assigned. The features are analyzed, using mainly the stereoscopic viewers; necessary measurements are made; and each feature is delineated on rectified imagery. As FADT information is determined--given feature identification codes and codes for the surface material--it is entered onto OPSCAN or keypunch FADT forms. Once all the features in a given manuscript area have been analyzed, the entire area's FADT forms are scanned or punched, and the data are read into the computer for validation. The verified data are then released for digitization on a standard coordinate grid. Prior to validation, all features are renumbered with a set of final ordered and unique FAC numbers.

A reduced sample of an OPSCAN FADT form is shown in Figure 6. The actual size of the form is 8-1/2 x 11 inches. The required handprinted numerics are shown as samples in the lower right-hand corner of the form. Note that all three horizontal lines in each matrix box are already drawn in, so the analyst need only enter the proper vertical strokes required for each

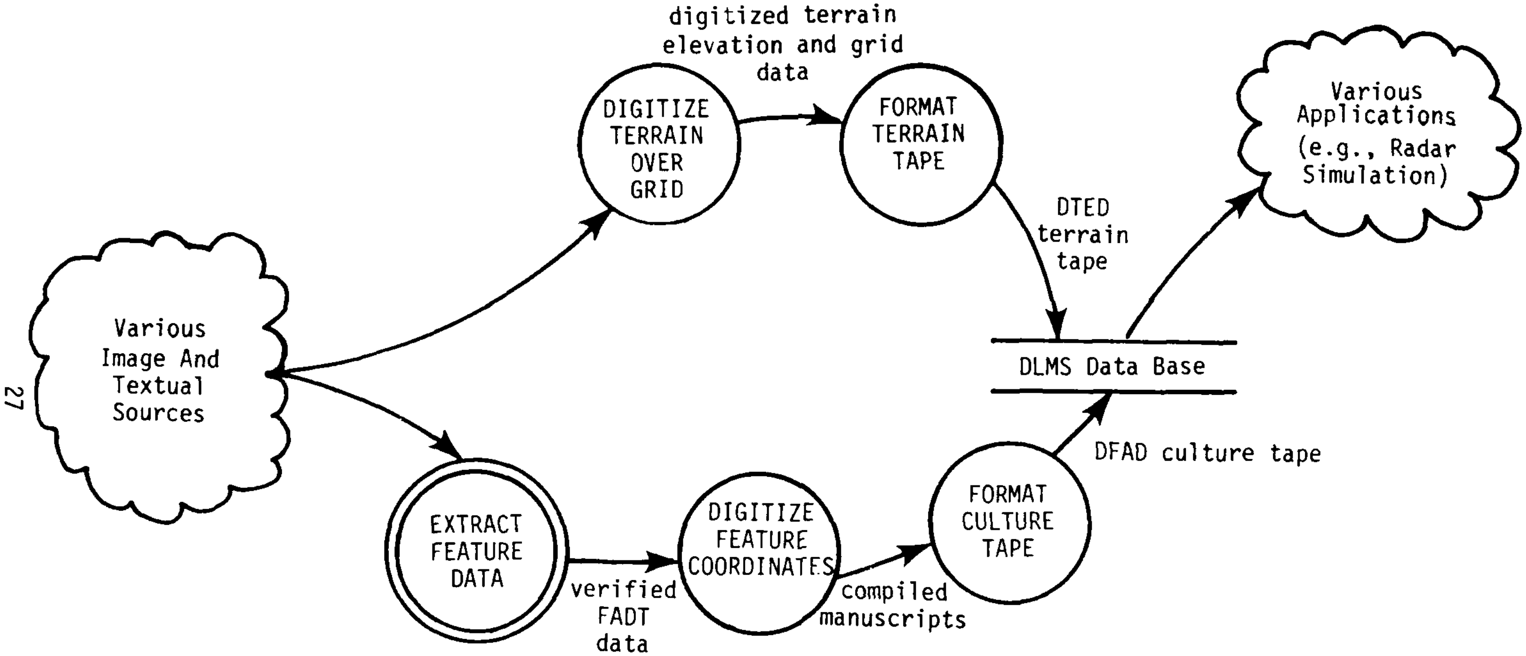


Figure 4. How the Digital Landmass System (DLMS) data base is created.

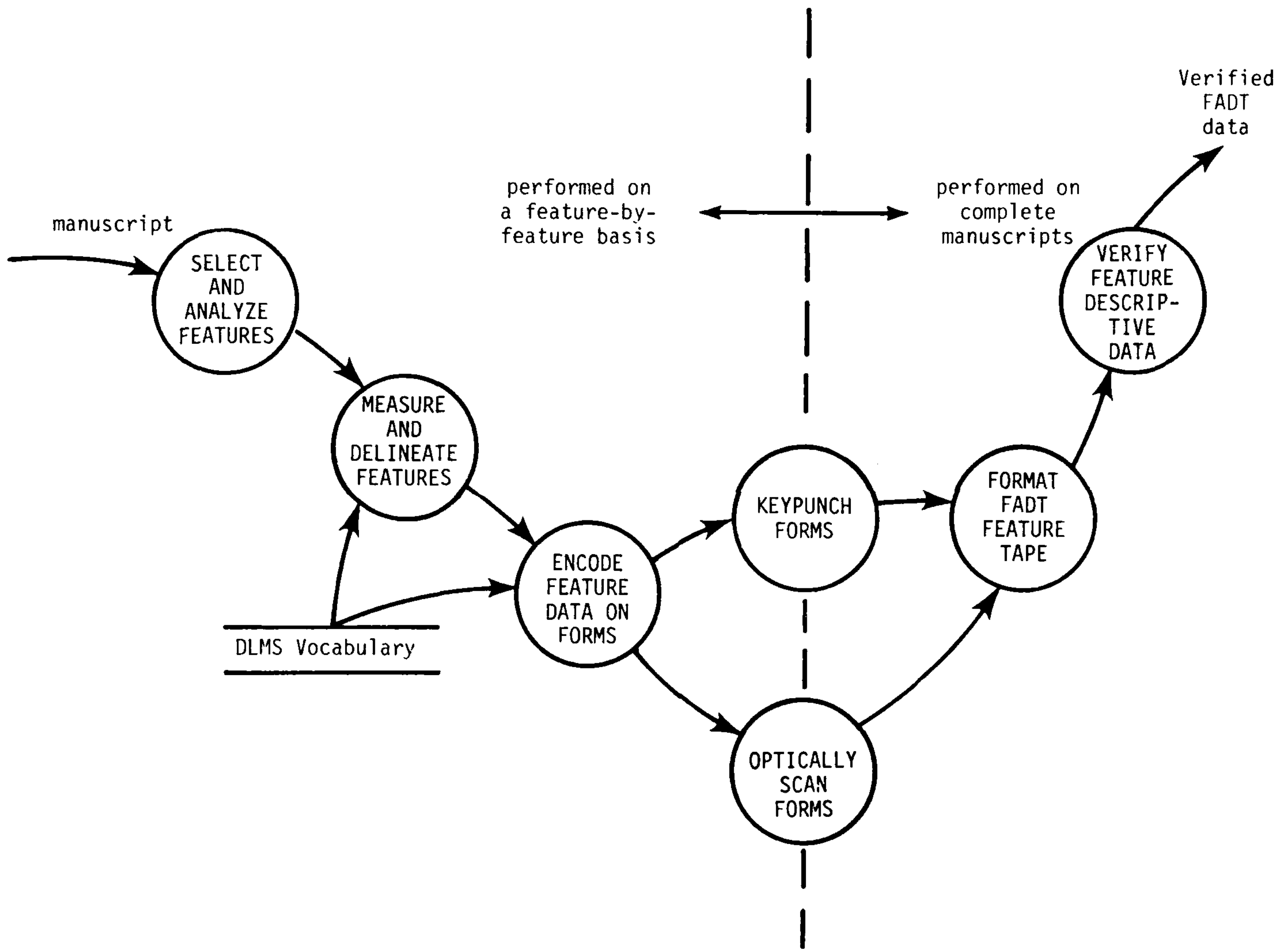


Figure 5. EXTRACT FEATURE DATA (present system).

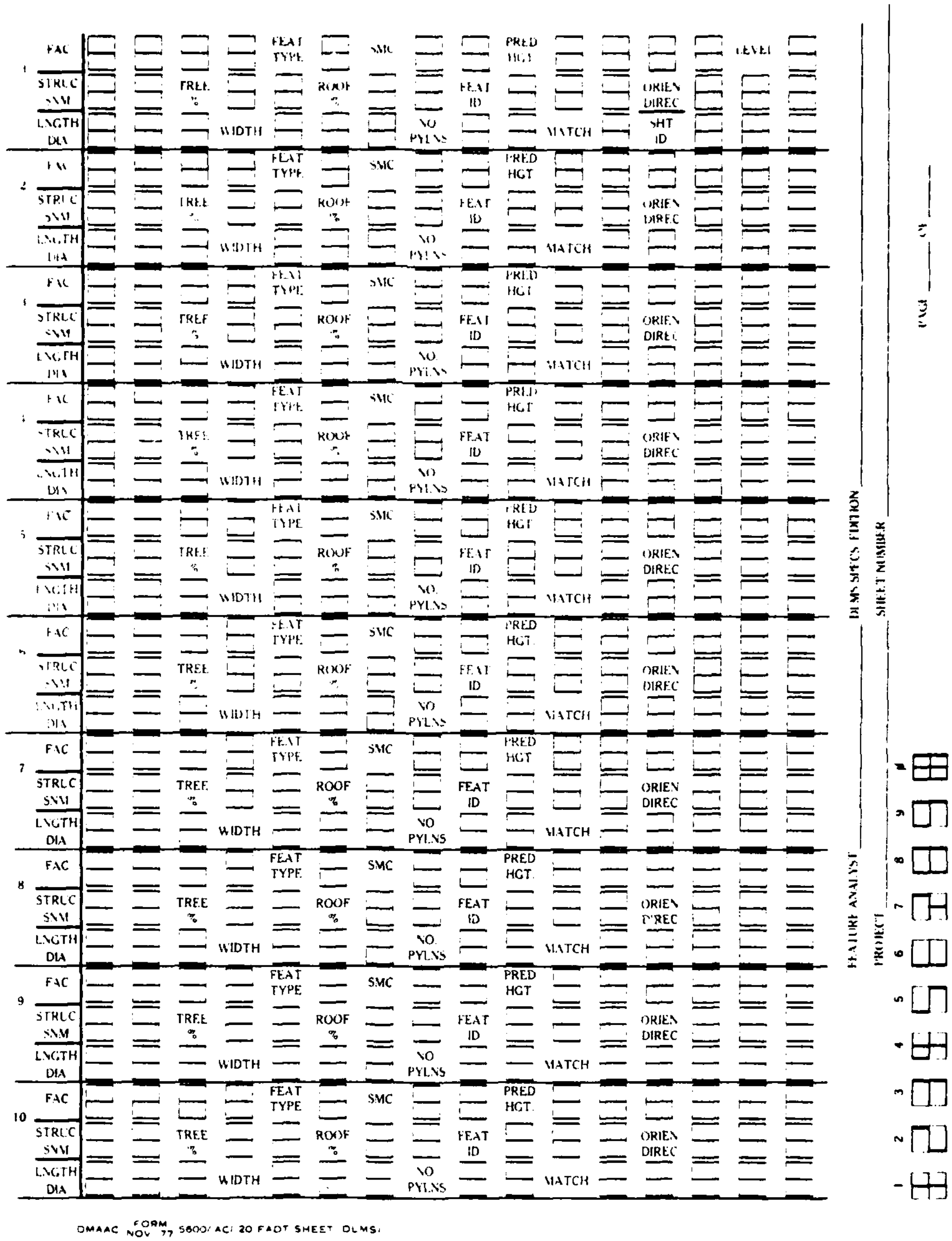


Figure 6. OPSCAN FADT data entry form (reduced in size).

number. Thus a "3" consists of a full-length vertical line on the right side of the box; the addition of a half-length vertical line on the upper half of the left side of the box would make the "3" a "9". (The number "6" is in error; the right vertical lower half of the box should also be filled in.)

The required numerics may be easily read by the optical scanner, but it is obvious that numbers are not normally written in one-eighth- or one-quarter-inch straight-line segments on a box consisting of a matrix of three vertical and three horizontal lines, the latter of which are always present. It can only be assumed that this unnatural and complicated method of hand-printing numbers results in far more errors than would a well-designed form geared more to the analyst than to the machine.

Information on data entry rates is essential to a multistation configuration study to determine where and when the system would become overloaded. A dearth of such information for FADT compilation exists, primarily because present data entry is off-line. Data entry rates have little meaning in an off-line process in which the only resources to be shared by analysts are pencils and FADT forms, the supply of which is essentially unlimited. A multiple-station configuration tied to a central computer, by contrast, requires that the computer's computing power, memory, and input/output (I/O) be shared. Such functions cannot always be divided easily or may not have the speed to respond quickly enough under high loads. Thus, if enough analysts happen to be entering data at the same time, the central processor could get overloaded and have to delay its responses to different analysts, slowing down the entire data entry process.

By knowing what the average and worst-case data entry rates are, the central processor can be designed with enough capability to meet some response-time criteria. One such criterion might be that the computer will respond within 1 second 60 percent of the time, and no response will ever be delayed more than 5 seconds at most. Once again, since present data entry is off-line, estimates of acceptable response times are mostly guesswork. From experience in other areas, however, it has been found that response times just a few seconds too long can result in user frustration and have significant negative influence on whether the system is accepted.

What data-rate information was available was provided by DMAAC, and these data, incorporated into Figure 5, are presented in Figure 7. Manuscript areas can range from 250 to 2,252 square nautical miles (snm), depending on the density of features in the area. On the average, there are over 1,500

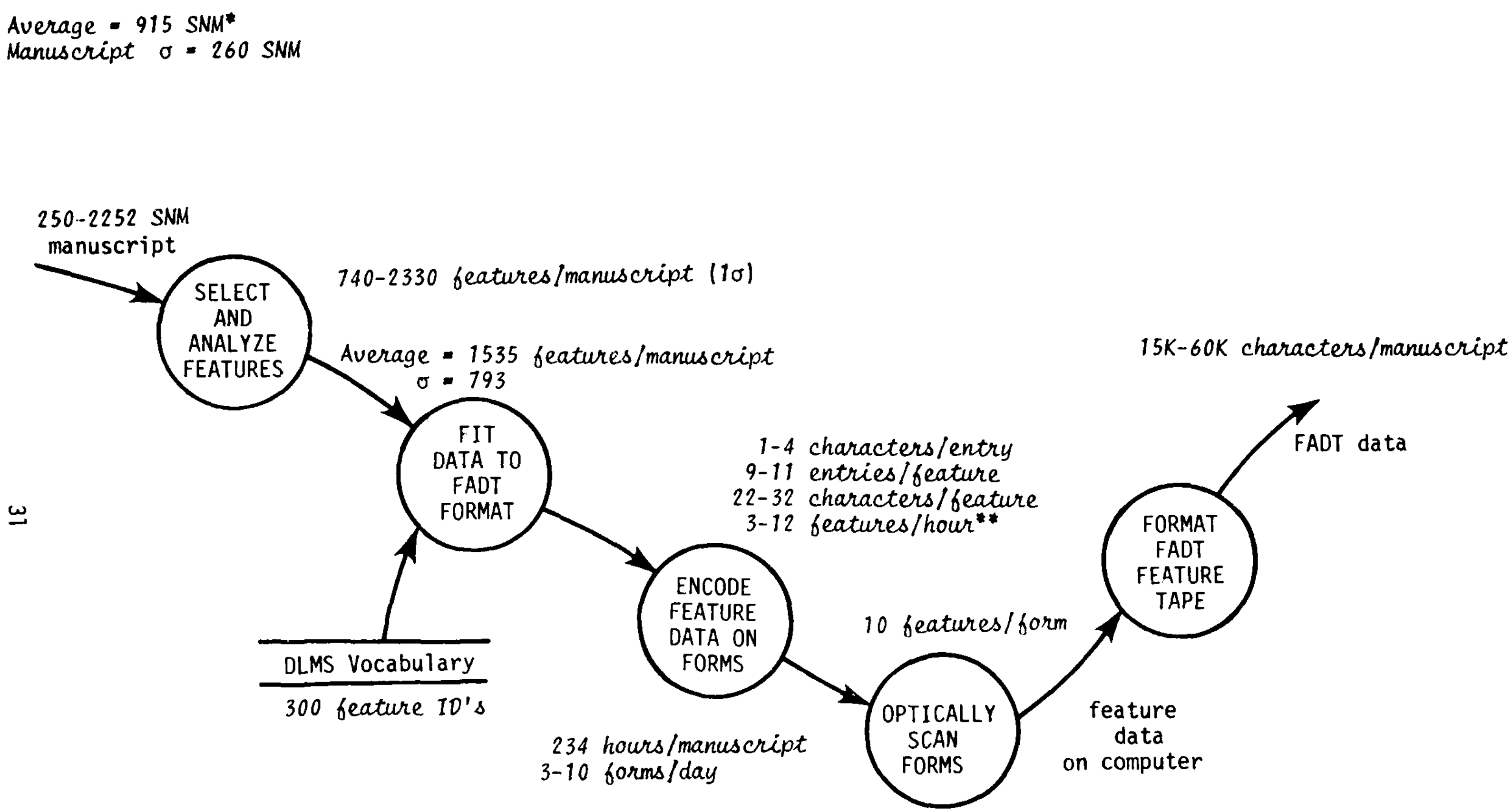


Figure 7. EXTRACT FEATURE DATA (present system).

*SNM = square nautical mile
**Average 6.5 features/hour, σ=3.7

features per manuscript, but with a standard deviation of almost 800. Between 8 and 11 entries are required for each feature, depending on whether the feature is a point, line, or areal feature. A typical point feature might be a water tower; a typical line feature might be a railroad track or a stream; and a typical areal feature might be a lake, a factory, or an airport.

Each entry will require from one to four characters (numbers) for its specification. For example, feature type requires only one digit, a zero, a one, or a two; but feature identification can be one of 260 numbers, and always requires three digits for its specification. In total, the entire feature could require from 22 to 35 numeric characters for its specification.

DMAAC estimated that it takes an average of almost 280 hours, or seven weeks, to complete a manuscript. Their standard deviation for this value was just over 120 hours, or three weeks. They indicated, however, that only six of the seven weeks, or about 85 percent of their time, was actually spent on tasks that require data entry.

This information yields an average data entry rate of just over six features per hour or one feature every ten minutes, which corresponds to about one entry per minute. In fact, DMAAC indicated that this figure could double during peak periods, or drop during periods when no data are entered. Thus the data entry rate can vary from essentially zero to one entry every 30 seconds. It must be emphasized that these figures are based on average off-line entry values and represent a wide range of individual-analyst procedures. Reliable on-line, real-time data entry rates will have to await experience with the proposed IFASS system.

It is interesting to compare the data entry rates derived above with those that appear in the ADM evaluation tests. For OPSCAN, voice, and keyboard data entry, CDI analysts averaged about three entries every minute. However, as was mentioned above, these two analysts were not performing analysis but merely specifying descriptive information on preselected features. Therefore, two entries per minute is probably a more realistic figure for the average peak data-entry rate in the present system.

2.2 VDE DATA FLOW REQUIREMENTS

A data flow diagram for voice data entry is shown in Figure 8. Adhering to statement-of-work requirements, the date flow includes voice and keyboard data entry; on-line, real-time training and retraining; audio and visual

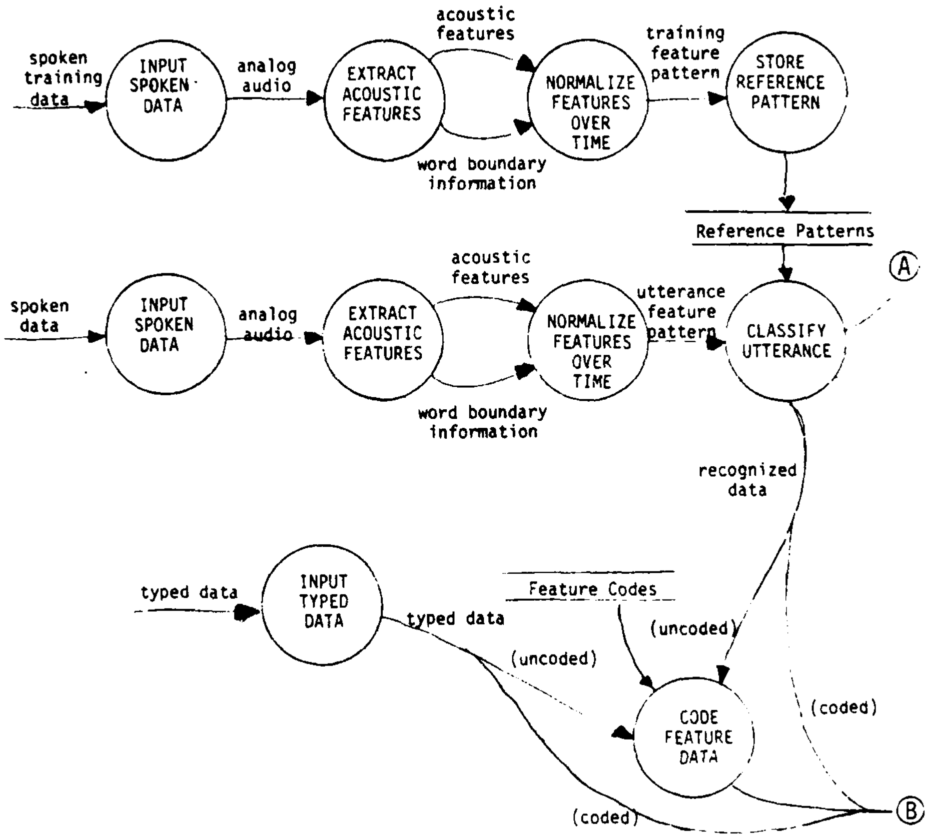


Figure 8. Data flow requirements for DLMS voice data entry system (p. 1 of 2).

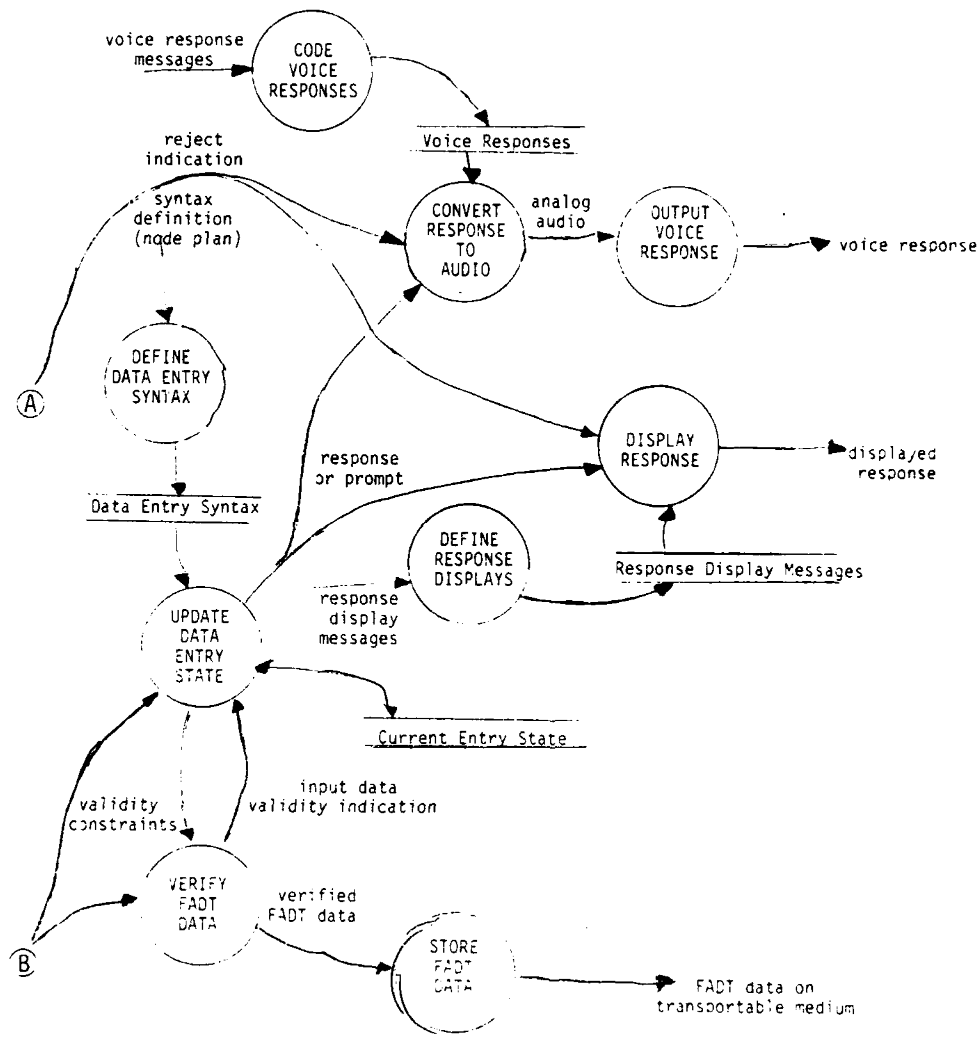


Figure 8--Continued (p. 2 of 2)

response to inputs; and storage of FADT data on a transportable medium in a DLMS-compatible format. Other statement-of-work requirements, such as a 500-word input vocabulary, speaker-dependent, isolated-word input with modular construction for future upgrading, and adequate safety requirements, are a function of specific hardware and system implementations.

The basic procedures for data entry are assumed to be similar to those used at present, except that data would be entered by voice instead of by hand. Since VDE is an on-line data entry process, various real-time response and verification functions can be added. Procedures that are specific to the use of <u>voice</u> as an on-line input--procedures for error correction and training/retraining--will also be added.

Several aspects of data flow in a VDE system are of particular interest because of their potential for increasing data entry rates. In the present system, it was estimated that the data entry rate varied from one entry a minute, on the average, to about two entries per minute during peak periods. However, if we define the amount of analysis time per entry as t_a and the actual data entry time as t_d, the following data entry schemes are possible:

Feature Analysis   Data Entry

Scheme 1: $|\longleftarrow t_{a1,2} \longrightarrow|\longleftarrow t_{d1,2} \longrightarrow| \ldots$

Analysis   Entry   Analysis   Entry

Scheme 2: $|\longleftarrow t_{a1} \rightarrow|\leftarrow t_{d1} \rightarrow|\leftarrow t_{a2} \longrightarrow|\longleftarrow t_{d2} \rightarrow| \ldots$

Total Entry Time = $\sum (t_a + t_d)$ for each feature

In the first scheme, each feature is analyzed and its corresponding FADT data are entered at one time. In the second scheme, the data are entered as soon as they are analyzed. In the present system, both schemes are used, depending on the analyst and the area being analyzed; but there is no difference in terms of average data entry times. In an on-line VDE system, however, which scheme is used will make a significant difference, as can be seen from the following diagram.

Visual Analysis: $|\longleftarrow t_a \longrightarrow|\longleftarrow t_a \longrightarrow|\longleftarrow t_a \longrightarrow| \ldots$

Vocal Entries: $|\leftarrow t_d \rightarrow|$ $|\leftarrow t_d \rightarrow|$ $|\leftarrow t_d \rightarrow|$

Total Entry Time = $\sum t_a$ for each feature

If voiced entries can take place simultaneously with analysis, the total time for entry would no longer be $\sum (t_a + t_d)$ but some smaller value, $\sum t_a$ being a minimum. Thus, if it takes five minutes to analyze a feature and one minute to enter all of its FADT information, the current scheme requires six minutes total for analysis and entry; but a VDE system might require only slightly more than the five minutes that it takes for analysis, since the FADT information would have been entered simultaneously--vocal entries could be made while visual analysis continued. Assuming that data entry takes only 20 percent of an analyst's time, then the on-going data entry possible with VDE should decrease data entry times by this same amount, or 20 percent.

The above is a simplistic calculation, however. Other factors are also involved in analysis that raise issues of time: How much of an analyst's time is devoted to concentrating on the feature? To what extent are his/her visual and mental concentrations broken by voiced entries as opposed to manual entries? Is it reasonable to assume that entries can be made simultaneously with analysis if FADT codes have to be looked up in the DLMS Specification? Will intense concentration change the analyst's voice to any significant extent? Do analysts have any preconceived notions about VDE and its usefulness?

Those and other questions were posed to analysts at DMAAC. Their responses are discussed in Section 4 and presented in detail in Appendix B. For our purposes here, two conclusions are of interest.

First, about one-half of the 84 analysts surveyed needed to refer to DLMS documentation fairly often, at least for feature identification codes (FICs) and areal feature codes. Analysts will, then, have to interrupt their analysis to look up these codes, regardless of whether they are entering data by voice, and it is not reasonable to assume that analysis will proceed unbroken with voice entry; FADT codes will still have to be looked up.

The VDE system could use actual feature descriptions rather than codes for vocal entries, and presumably the analysts would find it easier to remember descriptions than to remember triple-digit numeric codes. But when queried, only one-fourth of the analysts indicated that they would prefer to make all of their FIC entries this way. Another one-fourth indicated that they would actually prefer to make all of their FIC entries in the numeric codes. Since none of the analysts admitted that they had all of the FICs memorized, it can only be concluded that many of the analysts prefer looking up FICs, perhaps because they can then see all of the alternatives displayed before them. For example, an analyst could choose between four different types of cylindrical storage tanks, or five different types of radar antennas, or eight different types of bridges, without having to memorize codes or short descriptive names for each. Surely it would be very difficult to have to memorize either codes or descriptive names for 260 or more different features.

Second, only about two-thirds of the analysts enter all of a feature's FADT information as it is determined (scheme 2, shown above). This means that, for one-third of the analysts, VDE will not be a faster mode of data entry even if the analysts have all of the FADT codes or descriptions memorized. Analysis is performed similarly to scheme 1, with FADT data being entered for that feature all at once, after analysis has been completed, so being able to enter data by voice will not significantly affect data input time. Of course, increased data entry rates would provide incentive for analysts to change from an all-at-once-analysis-then-data-entry procedure.

In conclusion, data entry by voice could, theoretically, save whatever portion of an analyst's time is currently spent on data entry. However, for several practical reasons, these time savings would probably not be realized initially for roughly 35 percent of the analysts. In fact, no such time savings were apparent in the ADM evaluation tests using two CDI analysts in a simulated analysis task. Whether improved data entry rates become apparent with further use of the ADM VDE system remains to be seen. In any event, even if on-line VDE does not significantly improve data entry rates, it should provide advantages for FADT verification and data accuracy because it is on-line and interactive.

In a multiple-station configuration, system functions consist of recording FADT data, as well as communication with, and control of, each individual station. For DLMS, over 180 analysts are involved in various tasks related to DFAD compilation, and a useful configuration for production analysis work should include at least 42 analyst terminals or stations.

Analysts will be entering data asynchronously; it is therefore difficult to predict what actual processing and data entry rates would be. Assuming 100 analysts were simultaneously entering data at an average of one to two entries per minute at each station, processing requirements for FADT data could be as high as 100 bytes per second, a modest value. Further analysis of functional requirements, data entry and arrival rates, and storage requirements is presented in Sections 5, 6, and 7.

3. SUMMARY OF AVAILABLE VOICE DATA ENTRY (VDE) EQUIPMENT

This section describes commercial speech recognition and voice response devices, and headsets and wireless microphones that are compatible with them. For the purposes of this configuration study, only off-the-shelf equipment was considered. However, in view of the rapid progress being made in the state-of-the-art in speech recognition and speech synthesis and the advanced devices most manufacturers have in development, we have examined voice input and output devices as subsystem modules such that advanced capabilities can be considered as they become available in the coming years.

3.1 SPEECH RECOGNITION DEVICES

Speech recognition devices on the market today offer a wide range of capabilities, in packages varying from small, single-board units to large, rack-mounted devices. All commercial devices claim high recognition accuracies over various test vocabularies, but prices vary by more than an order of magnitude. This price variation is due to configuration and component differences: Some units consist of an acoustic preprocessor only; others include microprocessors for on-board recognition algorithms and vocabulary storage; still others include minicomputers and large-scale data storage devices.

3.1.1 *Types of Commercial Devices*

Some of the commercial speech recognition devices listed in Table 1, such as the Heuristics 5000/7000 Series, have only recently become available; others, such as Threshold Technology's VIP-100, have been available for almost a decade. Most of the equipment in Table 1 recognizes isolated words or short phrases; that is, any continuous utterance delineated by sufficiently long pauses (typically at least 200 milliseconds) becomes a recognizable entity. Thus "YES", "REPEAT", "CANCEL", "EARTHEN WORKS", "OBSERVATION TOWER", and "POLAR ICE PACK" are all valid "words" for a recognition vocabulary, provided that the normal pauses occurring within each "word" are not long enough to be mistaken for the end of the utterance.

In contrast to isolated-word devices are connected-word recognition systems, such as NEC's DP-100 or Dialog's Model 1800. These systems can recognize each individual word in a string of up to five words spoken

TABLE 1. COMMERCIAL SPEECH RECOGNITION DEVICES

Manufacturer	Model	Size	Weight	Power	Processor	Compatibility	Active Vocabulary Size	Price (Single Quantity)
Centigram Corp.	Mike	10" x 10" board Stand-alone unit	1 lb.	9 watts	DNA	RS-232C plug on optional board	16 words (up to 12 sets)	$1.8K - $4.5K
Verbex Corp.[1]	1810 117	Full-size rack units	500 lbs.	1,500 watts	DNA	Stand-Alone (RS-232C, Telephone, and other)	12* (out of 99 word choices)	[illegible]
Heuristics " " "	50,205 2,000 5,000 7,000	10" x 4.5" board 8" x 3.1" board 12.25" x 13.5" board 14.5" x 16.75" x 4.5"	DNA DNA 2.4 lbs. 11 lbs.	1 watt DNA DNA 50 watts	N/A N/A DNA DNA	S-100 bus Apple plug-in RS-232C plug "	64 words 32 or 64 words 64 words "	[illegible]
Interstate Electronics Corp.[2]	VRM VDES	6.75" x 12" board Several discrete units	DNA DNA	8.6 watts DNA	[illegible], 8KBRAM NOVA3, 64KBRAM	Intel Multi-bus and RS-232C Stand Alone (RS-232C, Telephone and other)	40, 70 or 100 words 250 to 850 words	$1.7K - $2.3K $19K - $44K
NEC America, Inc.	DP-100	Several discrete units	134 lbs.	DNA	DNA	Stand-Alone	120 words	[illegible]
Perception Technology Corp.	VE-100 VE-200	16" x 19" x 5" 15" x 19" x 10"	12 lbs. 65 lbs.	55 watts 550 watts	DNA DNA	PDP/? Stand Alone	14* + [illegible] 14* + [illegible]	[illegible]
Phonics[3]	SR/8R	5.5" x 10" board	DNA	19.2 watts	N/A	S-100 bus	16 words	[illegible]
Scott Instruments Corp.[4]	VET2	12" x 10" x 3.5"	5 lbs.	15 watts	N/A	TRS-80	40 words	[illegible]
Threshold Technology, Inc.[5]	VIP-100 TTI-500 TTI-600 TTI-580 TTI-680	Several discrete units	50 lbs.	125 watts	NOVA2, [illegible] 11, 8KBRAM NOVA3,	Stand Alone (RS-232C, Telephone, and other)	[illegible] to 500 words	[illegible]

*Speaker-Independent vocabulary requires no training

DNA - Data not available.

NOTES:
1. An affiliate of Exxon Enterprises, formerly known as Dialog Systems.
2. Interstate has recently announced three voice recognition chip sets, available in large OEM quantities only.
3. Phonics is now part of Interstate.
4. It has been reported that Texas Instruments has an option to purchase a nonexclusive license to use Scott's technology in TI products.
5. TTI has stated that they currently have a board level recognition device under development.

continuously . Thus a data entry such as "ONE, SEVEN, FIVE, TWO, SIX" could be made at once rather than as separate entries (i.e., "ONE", pause, "SEVEN", pause, ...), as required by the isolated-word systems.

Recently, Threshold Technology announced the introduction of their 580/680 QUIKTALK system, which attempts to bridge the gap between isolated-word and connected-speech recognition systems. Using data buffers to store several utterances, and an improved word-matching procedure, only very brief pauses between words are required by the QUIKTALK system. With this system, then, large strings of data can now be entered by voice almost as fast as they can be spoken.

With the exception of Dialog's Model 1800 (and Perception Technology's VE Series), all of the commercial speech recognition systems are <u>speaker-dependent</u>, meaning that each user must train the system to his/her own voice patterns for each word in the recognition vocabulary. Some systems, such as NEC's DP-100, require only two training samples, at most, of each word; others, such as Threshold Technology's devices, require ten training samples of each word. For Interstate Electronics' devices, only three training samples of each word are recommended for new users; seven are recommended for optimum recognition performance by experienced users.

Several single-board voice recognition devices are on the market today, as the table shows, in a variety of configurations. But all require a host processor or computer to control their speech recognition. These devices generally consist of a series of bandpass filters that isolate and store speech information within various frequency ranges over time. As each user trains the system for recognition of his/her voice, a group of stored patterns is created with which future utterances can be compared for recognition purposes.

3.1.2 Performance of Commercial Devices

The most critical requirement in any speech recognition system is recognition accuracy. More than any other factor, recognition accuracy determines the acceptability and usefulness of a voice input system. Most manufacturers quote accuracies for their devices of between 95 percent and 99 percent correct recognition; but recognition accuracy can vary greatly, depending on the particular vocabulary being used, the user, the ambient noise level, and other environmental factors. Many of the devices listed in Table 1 allow the user to set a rejection threshold such that, if a particular utterance is not similar enough to one of the prestored patterns, that utterance will be rejected rather than forcibly matched to the closest word. But, if the reject threshold is set too high, the normal variations in pronouncing a given word will result in false rejects and user frustration; if too low, spurious sounds or extraneous words could automatically trigger a recognition.

Unfortunately, some manufacturers will use high reject threshold to obtain just a few misrecognition errors at the expense of many rejects. In this way, a 95 percent recognition accuracy would be meaningless if one out of every four or five words was rejected as not being close enough to the stored word patterns. Even worse, an overly high reject rate could cause particularly difficult words to be repeatedly rejected--almost 100 percent of the time--bringing system operation to a halt.

To date, recognition accuracies for different devices have not been compared in any meaningful way by an objective third party. We present here some recognition accuracies reported by one manufacturer to show typical vocabularies and performance levels. Interstate Electronics has reported on recognition tests for their VRM board for various vocabularies, as follows: For a vocabulary consisting of just the ten digits, recognition accuracies varied from 99.4 to 100 percent for nine speakers, with an average accuracy of 99.9 percent; for a vocabulary consisting of 42 words (including the digits, the phonetic alphabet, and some common control words), recognition accuracies varied from 98.6 to 100 percent for the same nine speakers, with an average of 99.4 percent. Lastly, for a vocabulary of 100 words consisting of the previous sets plus over 50 phonetically similar (difficult-to-recognize) words, recognition accuracies varied from 90.3 to 98.8 percent for nine speakers, with an average accuracy of 95.8 percent. In all cases, tests were conducted using high-quality tape recordings of subjects in a quiet environment and using a zero-reject threshold (i.e., no rejects allowed).

In contrast to such generalized tests, a test of a 178-word vocabulary based on the DLMS specifications was conducted at Threshold Technology (TTI) and at RADC using the ADM (VIP-100) system. For the TTI personnel, recognition accuracies ranged from 94.4 to 99.3 percent for ten speakers, with an average accuracy of 98.0 percent. Speakers were only allowed to repeat a misrecognized or rejected word once before retraining; after retraining, the average accuracy rose to 98.7 percent. When these tests were repeated at RADC, ten speakers having different levels of experience had recognition accuracies which varied from 85.5 to 97.8 percent, with an average accuracy of 93.4 percent. After retraining, the average accuracy rose to 95.4 percent.

For the ADM evaluations tests conducted at DMAAC, it is difficult to determine recognition accuracy for the analysts tested because records of rejects were not kept. That is, only misrecognitions were recorded, and there seemed to be very few misrecognition errors. One of the recommendations of the ADM evaluation report was that continuous speech input and improved recognition accuracy would be advantageous.

3.2 VOICE RESPONSE DEVICES

Of the variety of speech synthesizers or voice response devices currently on the market, the newest are made up of only a few large-scale integrated circuit (IC) chips and offer unprecedented levels of efficiency and reliability.

Voice response systems include devices that store a fixed set of prerecorded messages on optical, magnetic, or electronic media, as well as true speech synthesizers, which transform a coded data string into a voiced message. Over a dozen manufacturers are involved with the prerecorded response systems, which tend to utilize older technology. These rack-sized units are generally designed for large-scale systems, such as telephone-based systems and communications with mainframe computers by multiple users. They incorporate small, fixed vocabularies designed for simultaneous use by as many as several hundred users.

Speech synthesizers come in three basic types: 1) those which use simple digital encoding of the speech waveform, 2) those which use some type of complex encoding of the speech signal, and 3) those which use linguistic or phoneme synthesis. A list of some commercial speech synthesis devices is provided in Table 2. Note that none of the devices listed uses simple speech encoding, which merely requires standard analog-to-digital conversion at a specified sampling rate, storage, and then a corresponding digital-to-analog synthesis for voice output. Such devices are application-specific--capable of being implemented in numerous ways--and hence no general-purpose product exists.

As Table 2 also shows, the synthesizer itself often consists of just a single chip (or part of a 2- or 3-chip set). Its incorporation into a voice response unit with some appropriate vocabulary and the required input/output

TABLE 2. A PARTIAL LIST OF COMMERCIAL SPEECH SYNTHESIS DEVICES[1]

Manufacturer	Model	Configuration	Type	Price
AI Cybernetic Systems	1000	10" x 5.5" board, S-100 bus compatible, 2.3 watts	Phoneme synthesizer	$0.4K
Computalker Consultants[2]	CT-1	10" x 5.5" board, S-100 bus compatible, 2.7 watts	Phoneme synthesizer	$0.4K
General Instrument Corp.	LISP-0256	Single-chip device providing 10 seconds of high-quality or 60 seconds of low-quality speech.	Complex encoding	$6
National Semiconductor Corp.[3]	SPC	2-chip set, 2K bits per second of speech	Complex encoding	$12
	DT 1000	Digitalker evaluation board, 5" x 6"	Complex encoding of 138 words	$0.5K
Speech Technology Corp.	M-188	10" x 5.5" board, S-100 bus compatible, 12 watts	Complex encoding of 64 words	$0.4K
	M410A/B	5.5" x 3.5" board which uses a single-chip synthesizer, RS-232C compatible	Complex encoding of up to 200 words	?
Telesensory Systems, Inc.	S-2B/C	3" x 3" board, < 1 watt	Complex encoding of up to 64 words	$0.1K-$0.2K
	?	2-chip synthesis set	Complex encoding	<$0.1K
	Series III	4" x 5" board	Complex encoding of up to 119 words	$0.4K
Texas Instruments	TMS-5100, 5200, 6100, 1000	1,3-chip set, 1.3K bits per second of speech	Complex encoding (LPC)	$15
	TM 990/306	Speech module	Complex encoding (LPC)	?
VOTRAX (Div. of Federal Screw Works)	VSB[4]	15" x 11", RS-232C compatible, 46 watts, 11 lbs (multi-lingual capability)	Phoneme synthesizer	$3.5K-$5K
	VSM-1	Versatile speech module, single board, RS-232C compatible	Phoneme synthesizer	$1.2K
	VS-6	12" x 11" x 3", RS-232C compatible, 30 watts, 11 lbs	Phoneme synthesizer	$3.5K-$5K
	CDS II	Custom development system for the SC-01 chip	Full-scale CRT terminal system	~$10K
	ML-IE	12" x 11" x 6", RS-232C compatible, < 46 watts, 20 lbs (multi-lingual capability)	Phoneme synthesizer	$6.7K-$8K
	PAC	Phoneme access controller, small size module	Phoneme synthesizer	$0.3K
	SC-01	Single-chip synthesizer, < 100 bits per second of speech	Phoneme synthesizer	$12K

NOTES: 1. Single-chip voice output devices have also been announced by ITT Semiconductors, Panasonic and Mitsubishi Electric Corp.
2. It was recently announced that this device has been taken off the market, awaiting new product development.
3. National has indicated that additional board-level synthesizers will be announced shortly, including a 150-word device designed for microcomputer systems, which uses National's BLX bus structure.
4. Also available as synthesizer-only (single board) units at quantity prices of $0.3K to $0.6K each.

interfaces is left up to the systems designer. Companies in Britain and France, as well as at least four major Japanese electronics firms, are also involved in developing single-chip synthesizers (or 2- or 3-chip sets).

These three synthesis techniques are described in detail in the following subsections to indicate the tradeoffs between the three types of speech synthesis methods. As Table 3 indicates, a general comparison of the three types of methods can be made by three key characteristics: quality of the output speech; memory (storage) requirements for output messages; and flexibility in the use of, and modifications to, output messages. To some extent, these qualities are interdependent, but for each criterion one of the three techniques is generally superior to the other two.

TABLE 3. A GENERAL COMPARISON OF THREE SPEECH SYNTHESIS METHODS

Synthesis Method	Speech Quality	Memory Requirements	Message Flexibility
Simple encoding	High	Greatest	Moderate
Complex encoding	Moderate	Moderate	Least
Phoneme synthesis	Low	Least	Greatest

3.2.1 Phoneme Synthesis of Speech

The key assets of phoneme synthesizers are an unlimited choice of vocabulary and minimal memory requirements; the chief drawback is poor-quality, machine-like speech.

Phonemes are the basic speech sounds of a language and correspond roughly to the individual letters in written language. In spoken English, over 40 phonemes are strung together by a complex series of interconnection rules to yield a connected utterance. In a phoneme synthesizer, these rules are approximated by additional phoneme sounds (called allophones) to account for sound variations with context, and a selectable set of durations, pauses, and inflections. Phoneme synthesis is often referred to as "synthesis by rule."

The VOTRAX division of Federal Screw Works has been a pioneer in phoneme-based speech synthesizers, their products ranging from a single-chip device to single boards and elaborate rack-mounted units capable of producing foreign languages. VOTRAX units are used for voice response by several manufacturers of speech recognition equipment, including Interstate Electronics and Threshold Technology, and are also being used to provide spoken output for IBM's talking typewriter. A block diagram of the VOTRAX technique for phoneme synthesis is provided in Figure 9.

In Table 4, data on three of VOTRAX's synthesizers are given. The user specifies (programs) those words and phrases to be spoken. Thus, vocabulary is unlimited , and messages can be generated or changed at will. In the case of their single-chip (SC-01) device, however, a $10K development system is required for the specification of vocabularies, in part because it often takes a great deal of phoneme and timing manipulation to make the speech reasonably intelligible. The minimal memory requirements of this technique, compared with those of other synthesis techniques, mean that vocabularies of many thousands of words are potentially feasible.

TABLE 4. PARAMETERS OF SOME PHONEME-BASED SPEECH SYNTHESIZERS MANUFACTURED BY VOTRAX

Model	Word Size	Speech Data Rate	Synthesis Functions
SC-01	6 bits	70-100 bps	45 phonemes, 16 durations, 3 pauses
VSB	8 bits	150 bps	61 phonemes, 3 pauses; plus 4 levels of inflection
ML-1	12 bits	300 bps	122 phonemes, 6 pauses; 8 levels of inflection (pitch) plus 4 phoneme durations

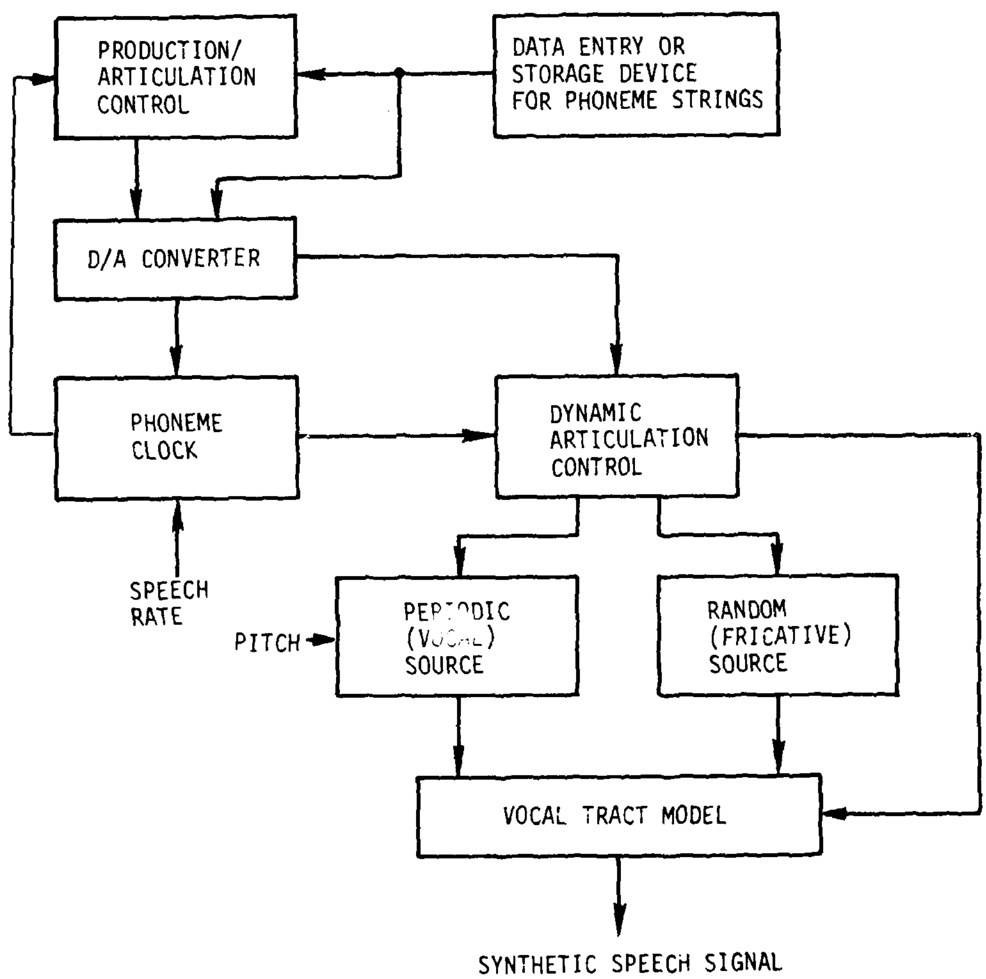


Figure 9. Block diagram of a VOTRAX phoneme synthesizer.

As shown in Table 4, data rates are as low as 70 bits per second (bps) of speech, which is roughly 1000 times below what direct digital recording of telephone-quality speech would yield. Considering this magnitude of data reduction, the quality of phoneme-synthesized speech is surprisingly intelligible under low ambient-noise conditions.

3.2.2 Complex Encoding of Speech

Synthesizers which use complex speech encoding offer improved speech quality over phoneme synthesizers, at the cost of increased memory requirements and decreased vocabulary flexibility.

Unlike the phoneme-based synthesizer, which can generate speech from text-like commands, other synthesizers use various compression techniques to encode a prerecorded audio message. The more complex the encoding scheme, the fewer the bits required for storing a word. The most common type of complex encoding used today is linear predictive coding (LPC), which typically uses a second-order digital lattice filter to model speech.

Texas Instruments was the first company to put LPC on a single chip. This chip was combined with a microprocessor-controller chip and memory chip for storage of word patterns to create a three-chip voice response unit. They were the first to develop an inexpensive commercial voice-response device, and this device gained wide publicity when it was marketed in an educational children's toy called Speak and Spell.

A block diagram of TI's technique for LPC synthesis is provided in Figure 10. The data rate for the TI chip set is about 1,200 bits per second of speech--several times that which can be obtained with phoneme synthesis, but still a considerable compression of the original speech waveform. Not surprisingly, the quality of speech resulting from LPC synthesis is significantly better than that obtained with phoneme synthesis.

The chief disadvantage of TI's LPC synthesizer, as well as that of other types of complex encoding schemes, is the necessity for prerecording and manipulating the desired utterances. Usually, someone with a professionally trained speaking voice is used to obtain samples of the required message, and often several hours of parameter manipulation by hand are necessary before the pattern is ready to be stored in memory for subsequent synthesis by the LPC chip. This technique is sometimes referred to as "synthesis by analysis," and one must typically go to the chip manufacturer to synthesize vocabulary. TI provides such a service at a charge of about $200 per second of speech.

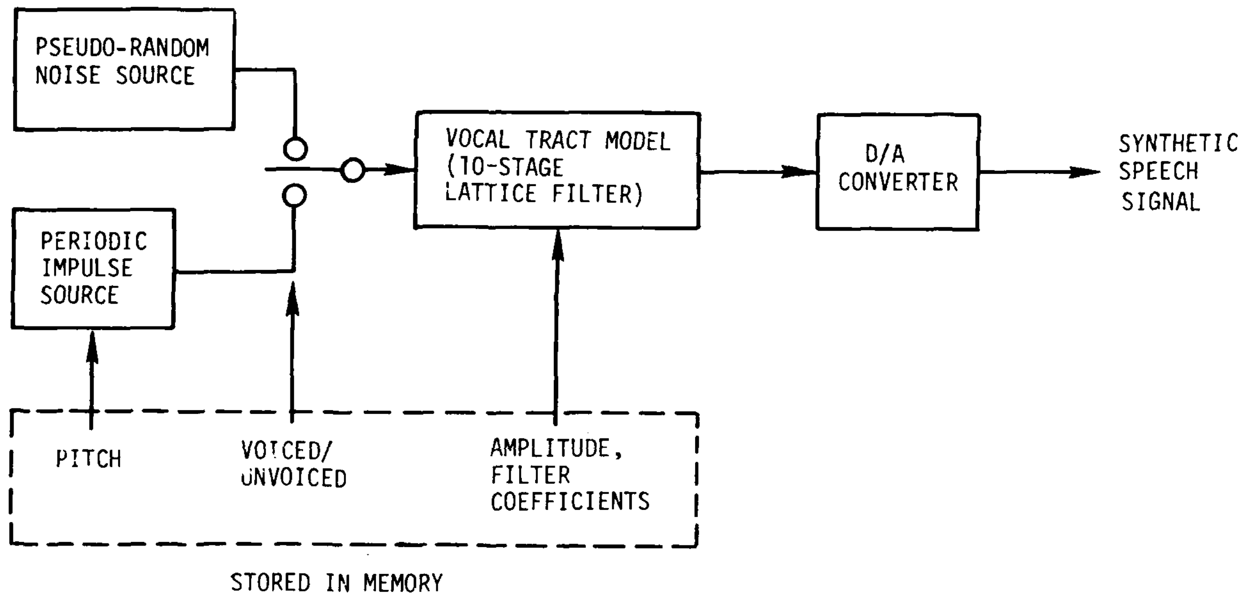


Figure 10. Block diagram of Texas Instruments' LPC synthesizer.

3.2.3 Simple Encoding of Speech

Techniques for simple speech encoding have the greatest memory requirements but offer the highest speech quality, along with moderate flexibility of vocabulary.

Simple encoding of speech using a straightforward compression technique such as adaptive-differential-pulse-code modulation (ADPCM), allows the direct storage of spoken messages in a manner analogous to a tape recorder, except that the speech samples are stored digitally on electronic media. Such speech-compression techniques are increasingly used to lower data rates in speech communication systems, making available single-chip units which do the A/D and D/A conversion, as well as the encoding, e.g., "CODEC" chips. While this technique may not have the glamour of other speech synthesis methods, it offers the highest speech quality.

3.3 HEADSETS/WIRELESS MICROPHONES

Several factors influence the selection of a headset for voice data entry:

- Comfort--The unit must be comfortable enough for long-term usage without causing undue fatigue.

- Sturdy--The unit must be rugged enough for everyday, production use.

- Noise cancelling--The sensitivity of voice recognition equipment requires that the microphone exclude extraneous noise.

- Stable position with respect to the inspector's mouth--This stability is required because the amplitude and quality of the sound field generated at the mouth vary significantly in space. Most noise-cancelling headsets require the microphone to be in close proximity with the mouth, which forces an analyst to swing the microphone's boom arm out of the way during stereoscopic viewing. Not only will such a procedure become tedious, but, unless the microphone can be repositioned accurately, recognition accuracy will decrease and possibly require retraining some of the vocabulary words.

A list of several noise-cancelling headsets for VDE at DMAAC is given in Table 5. All of the devices are sturdy, lightweight, and are used for telephone, aircraft, police, and similar professional applications. They also feature a single earpiece for proper feedback of the analyst's voiced entries. The frequency response for microphones and earphones gives an indication of fidelity, with all having at least 300-to-3,000-Hz minimum bandwidth.

The most attractive headset in Table 5, the Lear Siegler EarCom, is not a conventional headset, but a single-earpiece transceiver. The EarCom transmits speech by a transducer in the ear that detects voice energy via the otolaryngeal system. This same transducer is used as a standard earphone for receiving signals. With a custom-fitted earpiece, the EarCom provides very good comfort, stability, and noise cancellation, just as an earplug or hearing aid would.

Another physical concern, in addition to that of comfort, is possible interference with the headset's cord. As the analysts shift from their light tables to their stereoscopic viewers and their desks, the headset's cord could easily become caught or tangled. At the very least, this might knock the headset awry and require repositioning; at worst, such accidental entanglements might be painful if the headset is yanked completely off. A solution to this problem is using wireless transmission.

TABLE 5. SAMPLE HEADSETS

TYPE/MANUFACTURER/MODEL	WEIGHT	FREQUENCY RESPONSE MICRO PHONE (Hz)	FREQUENCY RESPONSE EAR PHONE (Hz)	COMFORT	STABILITY	NOISE
Earpiece Transceiver Lear Siegler EarCom 2683A	1oz[1] (28 gms)	300-3,000	300-3,000	Good	Good	Acoustic Attenuation
Professional Lightweight Shure SM-12	3oz (84gms)	50-15,000	70-12,000	Good	Poor	Electronic Cancellation
Professional Communications Telex CS-75	12oz[2] (340gms)	100-8,000	50 18,000	Poor	Good	Electronic Cancellation
Professional Lightweight Telex 5X5 Pro II	6oz[2] (170gms)	100-10,000	100-3,000	?	?	Electronic Cancellation
Lightweight Telephone UNEX HS-2A	1.4oz (45gms)[2] 7.5oz[2] (225gms)	100-5,000	150-3,000	?	?	Electronic Cancellation

[1]A belt-mounted pre-amplifier weighing 6 oz is not included.

[2]Gross weight for complete unit (includes cord).

Several tradeoffs must be considered before going to wireless communication. The chief advantage of wireless communication is that it provides the operator with complete mobility and freedom of movement. Its drawbacks are a function of the communication mode and equipment used. Four basic alternatives are listed in Table 6.

One-way audio transmission is the simplest method for achieving voice input. It offers no response capability, however, so prompts and data verification must be visual only.

Two-way, half-duplex audio transmission uses identical transceivers on a single carrier frequency for both transmission and reception on each end. This is a typical walkie-talkie type of system, and both the operator's unit and the system's unit are normally on standby. To activate transmission on either end, one transceiver must be switched to the transmit mode; the other transceiver automatically switches to the receive mode. This switching would be done electronically at the system's end to transmit prompts and verification of the operator's voiced inputs. At the operator's end a push-to-talk (PTT) or pressure-actuated microswitch would be used.

TABLE 6. ALTERNATIVES FOR WIRELESS COMMUNICATION

Wireless Mode	Hardware Components Required Per Station	Voice Transmission/Reception
1-way Audio	1 miniature microphone with RF amplification, 1 RF receiver	Voice input only (no response)
2-way Audio Half-Duplex	2 RF Transceivers (single channel)	Push-to-talk or other operator and system switching required
2-way Audio Full-Duplex	1 small RF transmitter, 1 RF receiver, 1 small RF receiver, 1 RF transmitter	Simultaneous voice input/output
2-way Infrared Full-Duplex	2 modulated IR transmitters 2 modulated IR receivers	Simultaneous voice input/output with no RF interference

The problem with such a half-duplex system is that the operator cannot transmit until the system is finished talking (and tests have shown that this will slow down the rate of voice inputs). More seriously, the operator blocks any feedback from the system, such as prompts, reject indications, error indications, or verifications, while he is in the transmit mode. That is, neither the operator nor the system can override each other, but either one can interrupt or block the other's transmissions inadvertently.

In a full-duplex system, two sets of transmitters and receivers are used. Each set is on its own carrier frequency, which means that transmission and reception by the operator or the system are independent and cannot be inadvertently blocked or interrupted. They can, in fact, take place simultaneously. A full-duplex system typically comprises a pocket-sized transmitter and receiver at the operator's end, with more bulky units at the system's end. The amount of equipment that must be attached to the operator and the additional weight he must carry around are thus minimized. While two sets of transceivers on different frequencies could also perform this function, such a system would not be as efficient or cost effective.

Depending on the range and the environment encountered, systems of a few tens of milliwatts to several watts might be required. Of course, output power affects not only the range and fidelity of the voice transmissions, but will affect the system's susceptibility to outside interference as well. Interference is, in fact, probably the major drawback to the use of wireless transmission for voice data entry systems. Needless to say, recognition accuracies will drop and system rejects will abound if other communications systems block or break into the communication channels being used. Also, the less powerful a given communication system, the more likely it is to be interrupted by other communication systems.

The FCC limits output power through licensing requirements and restricts the operating frequencies of even low-power transmitters to specific bands. Unless DMAAC has specific frequencies assigned to it, arbitrary frequencies would be assigned within the low-power industrial-use band. Whether these frequencies would be subject to frequent interruption would be entirely dependent on the local RF environment, and they might have to be modified on a trial-and-error basis. For a normal voice communications system, outside interference is merely a nuisance; in a wireless voice data entry system, however, outside interference would result in errors and poor overall system performance.

An alternative transmission mode that would eliminate RF interference would be an infrared light carrier for the voice signals. At least one manufacturer makes such a system with enough channels to allow full-duplex communication between four stations. But altnough an infrared system will not be affected by any RF communications systems, it does require direct line-of-sight or at least adequate indirect reflection between the transmitter and receiver. Thus, the larger the room, the more powerful the IR transmitter required, and under no circumstances could an operator in one room transmit to a voice entry system in another room. Depending on the workstation's configuration and layout within a particular area, this might or might not be advantageous.

Regardless of whether RF or IR transmission is used, the operator's transmitter and receiver units will have to be battery powered. Depending on the duty cycle of usage for a voice data entry task (percentage transmit, percentage receive, and percentage standby), batteries will usually have to be replaced or recharged daily. Long-life mercury or alkaline batteries or battery packs might yield one to two weeks' worth of use (assuming five 8-hour days per week), but the cost and nuisance of replacing or recharging batteries will remain a problem.

4. EVALUATION CRITERIA

Criteria for evaluating potential multistation voice data entry (VDE) configurations are listed below in the order of their importance.

1. User acceptance
2. Data entry accuracy
3. Data entry speed
4. Configuration flexibility
5. Reliability
6. Training requirements
7. Size requirements
8. Cost effectiveness
9. Development time and risk
10. Safety

The various factors to be considered when establishing these criteria are detailed in this section.

User acceptance. The VDE is designed as a man-machine system, to accommodate the habits and varying needs/desires of the human user. The degree to which the human user is willing to work with the machine--user acceptance--gives the most critical indication of system performance. A number of very subtle physical and psychological factors influence this acceptance. For example, operators must consciously speak in a clear and consistent manner for good recognition accuracies to be achieved. But most people's speaking habits are ingrained, and even the most cooperative speaker will be unable to change the habits of a lifetime overnight. No amount of coercion can force a speaker to cooperate in achieving good accuracy if he/she cannot accept the system. A headset that is uncomfortable to wear for more than an hour, or a display that is difficult to see, causes both physical and psychological fatigue to occur. And a system that takes too long to respond, or repeatedly rejects words, or must be retrained twice a day during hay-fever season, results in user irritation. The greater the user frustration, the worse system performance becomes and the lower the user faith in the system. Many of the following criteria also pertain to user acceptance.

Data entry accuracy. This criterion sets limits on how often errors can occur, how well they are caught in verification, and how easily they are corrected. Different types of errors are possible. For example, the analyst could remember the wrong FADT code, or speak a different code than the one he/she had meant to say, or mix up entries for one feature with those of another. None of these particular errors is caused by the machine, but the VDE system's error-correction capabilities must still be used. Of course, the most common VDE errors will be rejects or misrecognitions by the machine. It is hoped that a reject can be corrected merely by repeating the word a little more loudly or a little more carefully. In the case of misrecognitions--the machine recognizes a word different from the one that was spoken--correction will first depend on detection of the error from audio and/or visual feedback, then the machine's syntax being stepped back a node, the incorrect word being eliminated, and the correct word being spoken again. If the utterance is long, then dropouts may occur, meaning that part of the original utterance is missed and the resulting part-utterance is generally recognized incorrectly. If these types of errors occur frequently, user acceptance will drop and data entry will be slowed.

Data entry speed. The rate at which data are entered is affected by system response and verification times, as well as by error rates and error-correction times. Since pauses are required between words (i.e., isolated utterances), speaking rates will not approach those used in normal conversation, even if the system works perfectly (i.e., has no errors and instantaneous response times). The data rates that can be achieved will vary with VDE equipment, but even more so with user capabilities and experience. Thus, again, data entry speed and accuracy are closely tied to user acceptance.

Configuration flexibility. This criterion encompasses four separate areas: 1) the number of stations; 2) VDE capabilities; 3) vocabulary and syntax; and 4) other DMA application areas. The following paragraphs discuss each area in order.

It may be difficult and costly to add stations in a given configuration beyond a certain number. That is, a certain processor may adequately handle 16 user stations, but additional stations might require either a second processor or a significantly larger one. Several manufacturers offer standard multiterminal VDE configurations; for example, Interstate Electronics offer a four-terminal system that uses a Data General Nova 3 minicomputer. Custom configurations are also offered, and these are usually built up from standard configurations using a larger central computer. Lockheed Missiles and Space Company have one such multiterminal VDE installation consisting of 15 Threshold

Technology T-500 terminals tied to a Data General Eclipse S-130 minicomputer [Aviation Week 1980].* Depending on the central computer used in IFASS, it may be advantageous to use compatible equipment for VDE.

As to VDE capabilities, one consideration is ease of upgrading the system. Assuming that the speech recognition equipment chosen is speaker-dependent and requires isolated utterances, a question to be answered is, How difficult will it be to upgrade this equipment in the future as connected-speech and speaker-independent speech-recognition technology become more accessible and less expensive? The flexibility of the VDE equipment chosen for a particular configuration will affect the usefulness of the entire system for future upgrades.

The flexibility of the configuration to handle changes to the vocabulary and syntax of voiced entries is clearly spelled out in the statement of work. Such changes will also require the visual and voice-response prompts and verifications to be modified, and flexibility in these functions may be more difficult to provide. Ease in changing voice input or response vocabularies will allow for the substitution of words that are difficult for the system to correctly recognize (perhaps for only a few users), as well as vocabulary and syntax modifications required by changes to the DLMS product specifications.

This configuration study was specifically meant to address FADT data entry for DLMS compilation at DMAAC; however, the statement of work states that the configuration study should take into account other DMA application areas. A flexible vocabulary and syntax structure will go a long way toward attaining this goal, but other considerations pertinent to specific DMA applications must also be investigated.

Reliability. The potentially large number of stations in the multistation configuration presents the possibility of having a large number of electrical connections. Generally, the more connections, the lower the reliability. If all of these stations are, in fact, tied into a single central computer, any computer downtime will halt data entry unless each station terminal has enough intelligence and storage to continue recording information in the interim. The actual downtime, in that case, could be critical. The

---

*Aviation Week & Space Technology, "Word Recognition System Cuts Work in Parts Training," 15 September 1980, pp. 74-75 and 79.

solution to station reliance on the central computer is apt to be costly, regardless of whether intelligent terminals are used or redundant processors.

Training requirements. This criterion comprises individual voice-pattern training of each word in the vocabulary and training in the use of the system. Training in system operation will include procedures for system turn-on; analyst log-in; and downloading of programs and voice patterns, program execution, verification and error correction, data file manipulation, and logout. Ideally, training of voice patterns need only be done once; in practice, significant retraining will be required for new users and cases of illness or other physical or emotional changes which might alter voice characteristics. Some manufacturers, such as Interstate Electronics Corporation, recommend three training repetitions of each word initially until the user becomes familiar with the VDE system and his/her spoken tendencies and vocal control. The rationale is that new users are self-conscious and tend to pronounce words differently than when they are involved in day-to-day data entry. While additional training samples (Interstate recommends a total of seven each for optimum recognition accuracy) will improve recognition by a few percent, this advantage is lost unless the words are pronounced in a natural way.

Similarly, many VDE manufacturers recommend that training samples not be repeated for each word all at once, one after another, but that the entire vocabulary be cycled through the requisite number of times. That is, since people are not used to repeating a word over and over, for example, ONE, ONE, ONE, ONE,..., ONE, they should not have to train a vocabulary that way. Ideally, vocabulary words would be trained randomly; in practice, however, most VDE systems require that each word have the same number of training samples. A consecutive repetition of the vocabulary (e.g., ONE, TWO, THREE,..., ONE, TWO, THREE, etc.) is therefore a good compromise.

At any rate, ten repetitions of two or three hundred words could easily require an hour or more, which would be fatiguing. Theoretically only a single training session should be required; but training is an iterative process, and hence training and retraining requirements become an important system criterion as a function of the way the time consumed by training procedures affects user acceptance of the system.

Size requirements. Space on the analyst's workstation is already limited. One of the advantages of VDE is that a headset is all that should be required at the user's end; however, unless wireless communication is used, some type of user station is usually required. This user station should fit easily on the user's desk, and it will typically consist of an amplifier for the audio signal and some type of preprocessing or prefiltering equipment, as well as controls and displays for system initialization, feedback and retraining. Connected to each user station will be a headset for audio input and feedback, and perhaps a separate visual feedback display for use on the workstation (e.g., the LED display integrated into the stereoscopic viewers on the ADM system).

In a centralized, multistation configuration, space will also be required for the central processor and storage units, and for a main console for program operation, vocabulary definition, and syntax control. This main station would control and define program operation for all of the user stations.

Cost effectiveness. VDE systems vary in price from less than $1,000 to almost $100,000 for single-user stations. Much of this price variation is a result of different processing units and memory capabilities. For most stand-alone VDE systems, prices range from about $10,000 to $20,000 per station, depending on the specific equipment and hardware options. Other nonrecurring costs that must also be taken into account include the price of designing and implementing the hardware and software. In addition to hardware costs, hardware maintenance must be included as a recurring cost.

Development time and risk. The hardware described in Section 3 has been limited mainly to off-the-shelf equipment. It is also to DMA's advantage to consider recent developments in speech recognition and response hardware. Most of these developments are occurring at the chip level, however; so it will be a few years before full-up systems will reflect this progress. In particular, progress in single-chip microcomputers, speech synthesizers, bubble and semiconductor memories, and before long, speech recognizers, promises improved VDE performance at substantially lower costs. Nonetheless, these devices are just beginning to emerge today, and it may be advisable to await their future development.

Safety. Possible hazards associated with alternative configurations will be examined. For example, a configuration that has headsets attached by 6- or 8-foot cables to user stations that are, in turn, attached by 50- or 100-foot cables to some central computer may pose a significant safety problem if cables get snagged or people are tripped. Electrical safety problems may also exist.

5. VDE SYSTEM FUNCTIONAL DESCRIPTION

The functional description of a voice data entry (VDE) system for on-line compilation of DLMS feature analysis data bases is presented here. It is intended to establish a context within which the configuration analysis of the following sections can be examined. In Section 6, we define how the system will typically be used, and in Section 7, we make observations about three competing implementation approaches. From these observations, we develop quantitative results indicating the relative merit of the various configurations and complete the analysis by establishing an approximation of the performance that could be expected from the system (Section 7).

The following system is not to be construed as the optimal approach to the construction of DFAD data bases; we expect the IFASS system now under procurement by DMA to be eminently suitable to that application. Rather, our primary aim is to study the practical tradeoffs associated with voice input and output technology, free from the abstractions that arise when that technology is examined in a vacuum. In doing so, we have attempted to incorporate voice recognition and voice response capabilities into the application when those capabilities are appropriate, without overemphasizing their effectiveness as opposed to that of alternate methods of data entry and response.

In keeping with the above considerations, we have designed the VDE system primarily as an enhancement to manual data entry and visual response methods. We have subdivided system functions into the following three categories to reflect this subordination: 1) baseline functions requiring keyboard entry and CRT output only; 2) voice entry functions; and, finally, 3) voice output functions.

We preserve the groupings of functional capabilities to provide a basis for estimating the performance and cost associated with each. We hope that by approaching the analysis in this manner we can not only elicit a greater understanding of the implications of employing speech technology, but also provide a basis for judging the feasibility of adding these capabilities to the IFASS system at a later date.

5.1 SYSTEM OVERVIEW

A voice data entry system operates as an on-line data collection and storage device, providing an interactive interface between multiple users and two types of data sinks--hardcopy printout and magnetic tape storage. In the present case, analysts would work asynchronously and independently of one another to enter feature analysis data in the form of feature descriptor records. The system will automatically validate data entries as they are made, maintain storage for files of entries, and deliver these in the form of composite manuscripts on printed pages or portable tape reels. It is capable of providing these services to the entire analyst population concurrently.

Besides supporting real-time entry, the system offers comprehensive editing and review capabilities, procedural aids, computational support, and error reporting. Each analyst is provided with a functionally separate collection of resources: an entry station having two modes of data entry operation--keyboard and speaker-dependent automatic speech recognition--and two means of system response--CRT softcopy and synthesized speech. Analysts share access to the system printer, which enables them to receive listings of their own files.

The system operates largely unattended, although an operator position, served by a simple keyboard/CRT terminal, is provided. During normal operation the operator's position need only be manned to generate tape and hardcopy listings of composite manuscripts. The operator also maintains a list of authorized users and has control over modifications to the software configuration when necessary.

A view of the hierarchy of functions in the system is offered in Figure 11. At the first level, the system can be partitioned into on-line and off-line functions. On-line functions are those activities, initiated by individual analysts and by the operator, that relate to the day-to-day process of entering, verifying, and compiling feature analysis data. Off-line functions are those that are infrequently performed and are incidental to data gathering; general application program development and alterations to voice response vocabulary are examples of off-line activities.

On-line activities can be separated into those relating to the system operator and those relating to analysts as a group. Analyst functions are further subdivided into four major categories, as follows:

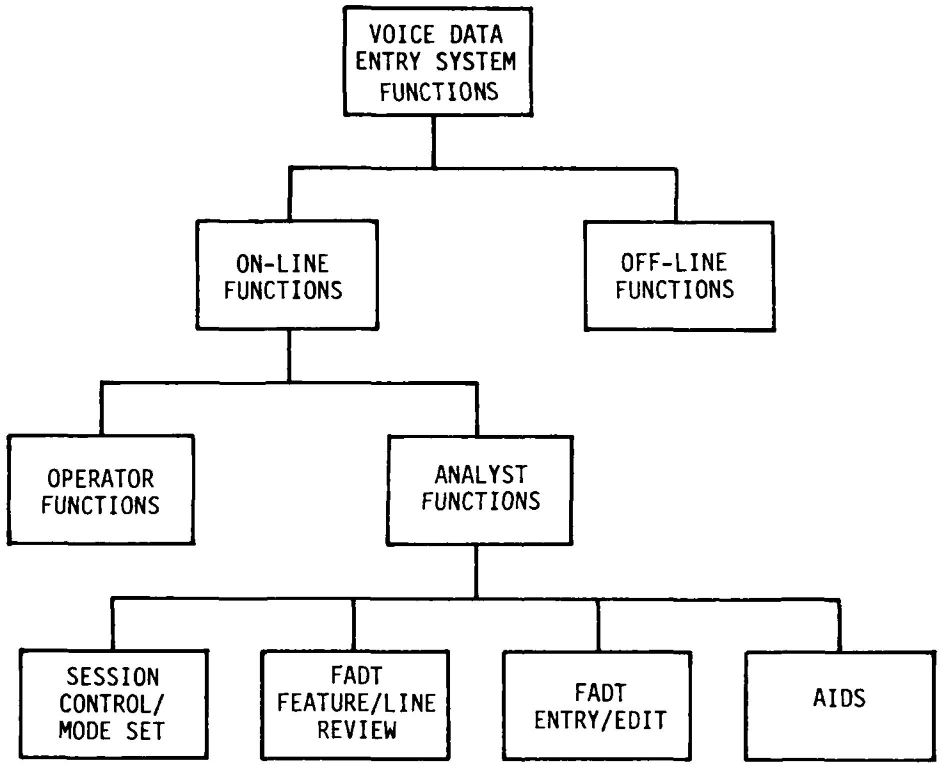


Figure 11. Overview of functions of data entry system.

1. Session control and mode setting
2. File review features
3. Data entry and editing
4. Informational and computational aids

In the remainder of this section we characterize the system functions in greater detail, focusing mainly on the four categories of analyst functions because their frequency of use is high and their influence on the performance of the system will be predominant. We spend less time on operator functions because their frequency of use will be much lower. We deal with off-line functions parenthetically, because we expect their frequency of use to be low enough that they will exert minimal influence. We will, however, attempt to account for their contribution in our later analysis (Section 7).

5.2 BASELINE SYSTEM FUNCTIONS: SAMPLE COMMANDS

The functions that define the baseline system require neither speech recognition nor speech response capabilities at the analyst interface. The analyst communicates with the system by manually entering alphanumeric data (and a small set of control characters); the system, in turn, returns alphanumeric data on a visual display. Thus, the interface required is no more sophisticated than a conventional keyboard/CRT terminal. Despite the relatively simplistic means of communicating with the analyst, the baseline system is by no means restricted in the services it offers.

In the description that follows, frequent reference will be made to commands entered at the keyboard. These suggested commands are merely examples of how the various functions could be implemented. An actual implementation would require a separate analysis of system software.

Commands take the form of a keyword that is followed, in some cases, by one or more parameters (signified by "{Parameter}") to be supplied by the analyst. Because the system can distinguish keywords by examining only a few leading characters, the analyst does not have to enter the entire keyword, except to enhance readability. Keywords will be shown as a concatenation of caps and lowercase letters; for the keyword to be interpreted correctly, only those characters in caps need to be typed. A summary of sample on-line commands for the baseline system is presented in Table 7.

All characters entered at the keyboard are mirrored by the system on the visual display, thereby providing the analyst a positive means of verifying that he/she has made them correctly. A cursor character is displayed

TABLE 7. SUMMARY OF SAMPLE ON-LINE COMMANDS, BASELINE SYSTEM

FUNCTION CATEGORY	COMMAND	ENTRY FORMAT
SESSION CONTROL AND MODE SETTING	LOGIN	LOGIn {Identification Code}
	OPEN FADT FILE	Open {Filename}
	CLOSE FADT FILE	CLose
	SUBMIT FADT FILE	Submit {Filename} {Manuscript Name}
	PRINT FADT FILE	PRInt {Filename}
	DELETE FADT FILE	DElete {Filename}
	LIST FILE DIRECTORY	LIstdir
	VERBOSE RESPONSE MODE	VErbose
	BRIEF RESPONSE MODE	Brief
	LOGOUT	LOGOut
LINE AND FEATURE REVIEW	REVIEW FEATURE DESCRIPTOR	FEature {FAC #}
	REVIEW FIRST FEATURE DESCRIPTOR	FIrst
	REVIEW LAST FEATURE DESCRIPTOR	LAst
	REVIEW TOP LINE OF FEATURE	TOp
	REVIEW BOTTOM LINE OF FEATURE	BOt
	REVIEW CURRENT LINE OF FEATURE	CUrrent
	REVIEW LINE ABOVE CURRENT LINE	Up
	REVIEW LINE BELOW CURRENT LINE	DOwn
ENTRY/EDIT	ENTER INPUT MODE	Input
	EXIT INPUT MODE	{Escape Character}
INFORMATIONAL AND COMPUTATIONAL AIDS	CONVERT CODE TO KEYWORD	COnvert {Code}
	CONVERT KEYWORD TO CODE	COnvert {Keyword}
	DELINEATE RANGE OF PARAMETER VALUES	RAnge
	DELINEATE FEATURE ID'S	RAnge {Feature Group Number}
	EVALUATE ARITHMETIC EXPRESSION	{Arithmetic Expression} =
OPERATOR	CHECK FADT FILE SUBMITTALS	Manuscript {Manuscript Name}
	DUMP MANUSCRIPT TO TAPE	DUMPTape {Manuscript Name}
	DUMP MANUSCRIPT TO PRINTER	DUMPPrint {Manuscript Name}

to mark the position of the next character to be input. The analyst makes entries one line at a time, signaling the end of each line in most cases with the carriage return character. If the analyst mistypes one or more characters, he/she can make a correction by backspacing one or more positions. Should the analyst wish to delete all the characters in the current line, he/she does so by typing "RUBOUT", which causes the cursor to backspace to the beginning of the line.

5.2.1 Session Control/Mode Set

Session control encompasses a group of functions that enable the analyst to independently gain access to the system and manipulate files being stored therein. Mode setting functions allow the analyst to select the method by which the system responds to his/her commands.

LOGIN

The LOGIN command is used by the analyst to identify him-/herself to the system to gain access to its services. The system bases its acceptance of the command on recognition of the analyst's identification code.

The command is entered on the keyboard in the form

LOGIn {Identification Code}

where {Identification Code} is a string of characters uniquely identifying the analyst. If the analyst's identification code is found in the list of authorized users (maintained by the operator), then the response "LOG IN SUCCESSFUL" is returned to the alphanumeric display; otherwise, a concise error message is displayed. It is not necessary for the analyst to be at an assigned station to use the system; any station will do.

Once the system has completed its log-in procedure, it notifies the analyst that it is ready for further inputs by displaying "RDY". The ready prompt is issued throughout the session whenever the system is capable of accepting a new command.

OPEN FADT FILE

Before the analyst can begin entering data, he/she must first access the particular file into which data will be entered. The OPEN command is used for that purpose.

The analyst enters the command in the form

Open {Filename}

where {Filename} is a character string specifying the name of the desired file. The system attempts to locate the file in bulk storage by consulting an internal directory. Within the directory, files are identified by both the filename and the identification code of the analyst who created the file. Thus an analyst might be allowed to access only those files bearing his/her identification code and only one file at a time, for example.

If the requested file is found in storage, the message "FILE OPENED" is returned to the display. If no file bearing the given name and identification code is found, the system assumes the analyst is requesting a new file file to be created. After storage for the new file has been allocated and its name and the current date entered in the directory, the message "NEW FILE CREATED" is displayed. Once the analyst receives the ready prompt, he/she is then free to review or edit the file.

CLOSE FADT FILE

When the analyst has finished with the file he/she is working on, he/she must close that file before moving on to another, by typing "CLose". The system responds with a message in the form

FILE {Filename} CLOSED ,

where {Filename} is the name of the file (as before), repeated as a reminder. If the analyst attempts to use the command inappropriately (such as when no file has previously been opened), an error message is returned.

SUBMIT FADT FILE FOR OFF-LINE STORAGE

When the analyst has completed the process of entering data into a file, he/she uses the SUBMIT command to copy the file to tape storage and to automatically post a notification of its completion for the operator.

The command is entered at the keyboard in the form

Submit {Filename} {Manuscript Name}

where {Filename} and {Manuscript Name} are character strings denoting the names of the designated file and the manuscript of which the file is a part, respectively. The system checks the named file to verify that there are no incomplete or invalidly specified features. If discrepancies are found, an error message is returned, and further processing of that file is aborted. Otherwise, a copy of the file is placed in temporary storage, and the identity of the file is entered into a list of files (associated with the same manuscript) that are awaiting transfer to tape. The system acknowledges receipt of the command by displaying a message of the following form on the analyst's display:

FILE {Filename} SUBMITTED TO {Manuscript Name}

where {Filename} and {Manuscript Name} are as before.

PRINT FADT FILE

When the analyst needs hardcopy of a particular file, he/she can employ the PRINT command to obtain it. The format of the PRINT command is as follows:

PRInt {Filename} .

The system acts on the PRINT command by making a copy of the file and internally queueing a print request. The system responds to the analyst by displaying the message "PRINT REQUEST IN QUEUE". When the printer becomes available after servicing requests ahead of the current one, the file is printed, and storage for the copied file is released for reuse.

DELETE FADT FILE

Files that are no longer needed by the analyst can be deleted, thus freeing internal storage for reuse. The DELETE command is used for this purpose. The command is entered from the keyboard as follows:

DElete {Filename} .

The system responds by deleting the specified file and returning the following message to the analyst:

FILE {Filename} DELETED

LIST FADT FILE DIRECTORY

The analyst may have a number of FADT files stored in the system at any time. Some files may be finished, and others may still be in the process of being compiled. The analyst can obtain a listing of all the files stored in the system under his/her identification code by typing the command "LIstdir".

The system responds by displaying a table of the form:

FILENAME -- CREATION DATE -- DATE OF LATEST MODIFICATION

Filename	Creation Date	Last Change Date
.	.	.
.	.	.
.	.	.

where the second and following lines (if any) list the name of each file, the date on which the file was created, and the date of the latest change to the contents of the file. If the analyst's file directory contains no files, the message "DIRECTORY EMPTY" is returned, instead.

System Response Modes

The system employs two modes--Verbose and Brief--of formulating responses directed to the display while the analyst is engaged in entering feature data. The Brief Response mode is entered by default when the analyst completes the log-in procedure. In Brief Response mode, acknowledgments sent to the analyst during data entry and file review tend to be more succinct than in Verbose Response mode. In the Verbose Response mode, for instance, feature identification codes entered by the analyst are converted to their equivalent English keywords and displayed as a means of verifying that the analyst has entered the appropriate code. Likewise, if the analyst chooses to enter feature analysis keywords rather than feature identification codes, the corresponding code is displayed.

To enter the Verbose Response mode, the analyst types, "VErbose" at the keyboard, and the system responds with "VERBOSE MODE". To reenter the Brief Response mode, he/she types "BRief", and the system replies with "BRIEF MODE". Analysts who become readily familiar with codes and keywords will probably choose to receive brief responses, because the system's response time will be somewhat improved and the display will be cluttered with less unnecessary information. Other will wish to take advantage of the expanded responses to save having to manually crosscheck and verify codes and keywords.

LOGOUT

When the analyst has finished using the system, he/she logs out by typing the command "LOGOut". The system responds with the message "LOG OUT ACCEPTED". Once the analyst has logged out, the system ignores all other commands except LOGIN. Thus, the system can be protected against unauthorized access during periods when the analyst's station is unattended.

5.2.2 Review Functions

Review functions are those that allow the analyst to ascertain what information has been entered into an FADT file. The system organizes FADT files as a linear collection of feature descriptors ordered by feature analysis code (FAC) numbers. FAC numbers begin at 1 and increase sequentially. Entries within each feature descriptor are arranged into discrete lines, each line corresponding to one of the 13 columns, A through L, shown in the feature analysis data table of Figure A-1, Appendix A.

Feature Review

Three commands are provided the analyst for reviewing a feature descriptor in its entirety. If the analyst wishes to review a feature whose FAC number is known to him/her, he/she enters a command of the following form in the keyboard:

FEature {Feature Analysis Code Number}

where {Feature Analysis Code Number} is a number from 1 to 9999.

If there is no feature corresponding to the given number in the file, the system returns the error message "FEATURE NOT FOUND". Otherwise, the system displays the feature descriptor in tabular form on the alphanumeric display. In the Verbose Response mode, a feature appears as in Table 8. Here, the columns of Figure A-1 are translated into rows, each containing a heading in English, followed by its entry, if any. Entries are shown in coded form, followed by their corresponding English keyword, where applicable. In the Brief Response mode the display is similar, except that neither keywords nor the units of measurement for various entries are displayed.

The analyst can review the first feature in the file by typing the command "FIrst" at the keyboard. Likewise, by typing the command "LAst" he/she can review the last feature entered.

Line Review

The analyst has five commands to aid in reviewing a single line of a feature descriptor, if desired. The system internally maintains a current line pointer as an index to the open FADT file. When the analyst opens a file, the line pointer is set by default to point to a line within the last feature descriptor of the file. The line it points to is the one containing the last entry in the feature (relative to the feature descriptor format described in Table 8). When a file is newly created, the system automatically begins to fill out the first feature descriptor by setting the "FAC #" field to a value of 1 (which the analyst is free to change if he/she wishes). Thus, in an empty field, the line pointer points to the "FAC #" line of the first feature descriptor.

The analyst can position the line pointer at the top or bottom of the current feature description by entering the command "TOp" or "BOt", respectively, and can also move the pointer up or down one line at a time by keying in "Up" or "DOwn", respectively. Using the UP and DOWN commands, he/she can step the line pointer across a feature descriptor's boundary into the adjacent feature. For instance, if the line pointer is at the bottom of a given feature, the DOWN command will move it to the top line of the following feature (once there, the TOP and BOTTOM commands are interpreted as applying to that feature and not the previous one).

TABLE 8. SAMPLE DISPLAY FORMAT FOR FEATURE REVIEW (VERBOSE MODE)

FAC #: 1Ø4
Feature Type: Ø (Point)
Surface Material: 3 (Stone/Brick)
Predominant Height: 1Ø Meters
Structures Per Sq. N. M.:
Percent Trees:
Percent Roof:
Feature Ident: 65Ø (Church)
Directivity/Orientation: 4 (45 Degrees)
Length or Diameter of Point Features: 3Ø Meters
Width of Line or Point Features: 3Ø Meters
Level:
Number Pylons:

In all of these cases, the system responds to the commands by displaying the line pointed to, in the format depicted by Table 8 (i.e., the line heading followed by a colon and the entry, if any, recorded for that line. If the analyst wishes to review the line currently being pointed to, he/she can do so by entering the command "CUrrent". Whenever one of the feature review commands is used, the line pointer is automatically set to the line containing the last entry in the addressed feature.

5.2.3 Entry/Edit Functions

Entry and edit functions enable the analyst to add feature descriptors to an FADT file on a parameter-by-parameter basis and to modify data there as necessary. During the process of entering and modifying data, the system performs validity checks to verify that the individual entries in each feature descriptor are within allowable ranges and consistent with each other. As an aid to configuration control, the system automatically keeps track of the date on which the file was last revised.

Input Mode

Before new data can be entered into a feature descriptor, or old data modified, the feature and line review commands discussed above may need to be employed to place the line pointer at the desired line within the selected feature. As a prelude to entering data, the analyst must place the system in the Input mode by typing the command, "Input".

The system responds by displaying the current line, heading first, followed by a colon and its current entry, if any. Note that headings within a feature descriptor are never entered by the analyst; they are a permanent part of each feature descriptor. Empty feature descriptors are replicated automatically whenever the line pointer advances beyond the last line of the preceding feature descriptor.

If the analyst desires to change the value of a descriptor parameter or make a new entry in an empty parameter field, he/she does so by typing the correct value and terminating the entry with a carriage return. Should the entry be mistyped, the analyst can always backspace or delete the entire line by typing "RUBOUT". If the analyst decides not to alter the current entry (or leave the field blank until later), he/she types carriage return alone. As an aid to configuration control, the system keeps track of the date of latest revision.

In making entries into the feature descriptor, the analyst has the option of entering either numeric codes or feature analysis keywords for the following fields:

1. Feature type
2. Surface material
3. Feature identification
4. Directivity

The keyword options for each of these fields are indicated in Appendix A. Feature analysis keywords, as opposed to command keywords, must be entered in their entirety.

Each entry received by the system is verified to be a recognizable numeric code or keyword. Keywords are converted into corresponding codes. The coded entry is validated by being compared against the range of allowable responses for the current descriptor field. Lastly, the entry is verified as being consistent with all parameters that have been previously entered into that feature descriptor. If the entry fails any of these checks, an appropriate error message ("UNRECOGNIZABLE ENTRY", "INVALID ENTRY", "INCONSISTENT ENTRY") is displayed to the analyst. Additional information could also be provided as to specifically what the recognized error appeared to be. The system then waits for a corrected entry for the same descriptor field.

An additional level of verification is employed when the Verbose Response mode has been selected by the analyst. For descriptor fields in which either numeric codes or keywords are accepted, the system displays the alternate representation for each entry made by the analyst. The translation allows verification of the correct code, if the analyst takes advantage of the abbreviated entry form provided by the codes; similarly, if he/she enters keywords, then seeing the corresponding code will tend to reinforce the analyst's memory for that code in the future.

Once the system accepts an entry, it moves automatically to the next descriptor field and, as the last line in a feature descriptor is entered, to the next feature descriptor. The system automatically fills in entries for two descriptor fields--"FAC #" and "LEVEL"--during the entry process. The "FAC #" entry is selected by the system to be equal to 1 plus the value contained in the corresponding field of the previous feature descriptor. The analyst has the option of changing this value if he/she desires to initialize or to alter the sequence of FAC numbers. For the "LEVEL" field, the system copies the entry for that field from the previous feature descriptor. Thus, the analyst is freed from making entries in these two fields once he/she sets their values in the first feature descriptor record of a file.

When finished with making entries, the analyst quits the Input mode by typing the "ESCAPE" character, to which the system responds with the "RDY" prompt. The system is now capable of accepting all commands.

5.2.4 Informational and Computational Aids

AIDs are a supplementary service extended to the analyst to assist him or her in the process of entering data. Informational assistance is provided to free the analyst from the necessity of contantly relying on supporting documentation, such as tables of feature analysis codes. Computational assistance is offered so that a separate calculator will not have to be used in computing feature dimensions.

Code/Keyword Conversion

The analyst can ask the system to convert a given numeric code into its corresponding English-language equivalent keyword by entering the CONVERT command in the following form:

COnvert {Code}

where {Code} is a digit string having one to three characters. If the code is valid, the system will respond by displaying the corresponding feature analysis keyword. Otherwise, the error message "INVALID ENTRY" is returned.

Should the analyst wish, instead, to obtain the numeric code that corresponds to a given keyword, he/she enters an alternate form of the command:

COnvert {Keyword}

where {Keyword} is a character string. In this case, the system responds with the corresponding feature analysis code.

Parameter Range Information

When the analyst is unclear about the range of acceptable values for a particular field of the feature descriptor, he/she can use the RANGE command to request assistance from the system. The simplest form of the command is entered from the keyboard as "RAnge". The system interprets the command as referring to the descriptor field now being pointed to by the line pointer. It is assumed that an FADT file is open at the time the command is given; if not, an error message is returned. The system responds by displaying one or more lines of tabularized information describing acceptable parameter bounds, keywords, and corresponding codes.

This feature completely describes permissable ranges for all descriptor fields other than the "Feature Ident" field, which has too many possible values to be shown at the same time on a conventional CRT display. In this case, the system displays the nine major groups (100 through 900) of feature identification codes and their English titles. The analyst can then use an alternate form of the RANGE command to obtain a complete listing of the allowable values within any one of those groups by entering the command in the form

RAnge {Feature Group Number}

where {Feature Group Number} is one of the three-digit strings "100", "200", "300", ..., "900".

Arithmetic Expressions

The system provides the services of a scientific calculator to assist the analyst with dimension computations. The analyst uses this feature by typing an arithmetic expression and terminating it with the character "=" rather than the carriage return. The system responds by displaying the answer in scientific form (mantissa and exponent).

Arithmetic capabilities of the system (and the operators used to specify them) include addition "+", subtraction "-", multiplication "x", division "/", and exponentiation "**". Trigonometric forms, sine ("sin"), cosine ("cos"), and tangent ("tan") are also available. Arithmetic expressions are entered by interspersing arithmetic operators and real numbers on a single line without intervening spaces. Parentheses ["(" and ")"] may be used in the conventional manner to specify the precedence of arithmetic operations.

5.2.5 Operator Functions

The operator directs the transferring of manuscript files to off-line magnetic tape storage and the printing out of hardcopy. These procedures occur only once a day, at which time the operator must mount and dismount tapes and maintain the supply of paper in the printer. Since these activities involve only a few minutes during the day, a separate full-time operator is not necessary.

The operator position makes use of a keyboard/CRT terminal to interact with the system. Its capabilities parallel those of the analyst position in the baseline system configuration. In addition to being able to enter the privileged commands that can be exercised only from this position, the operator has access to all the commands that have been defined in the baseline system for use by the analyst. The operator position affords a command interface for the operator, and, in addition, all system error messages are directed there.

CHECK FILE SUBMITTALS

The system maintains a list of FADT files that have been submitted to each manuscript currently being compiled. The operator can review the list to determine if any files remain to be submitted, by entering the command on the keyboard in the form

Manuscript {Manuscript Name}

where {Manuscript Name} is a character string identifying the manuscript of interest.

If no files have been submitted, the system responds with the message "NO FILES SUBMITTED"; otherwise, a list of submitted files is displayed in the same format as that used to display an FADT file directory (see Subsection 5.2.1).

Once the manuscript is transferred to off-line storage, the operator can delete the master copy of the manuscript and the submittals list by exercising the DELETE command of Subsection 5.2.1, with the name of the manuscript specified as the filename.

DUMP MANUSCRIPT TO TAPE STORAGE

When all outstanding files have been submitted, the operator can dump the completed manuscript to either tape or hardcopy, as he/she desires. A tape dump is achieved by entering the command

DUMPTape {Manuscript Name}

where {Manuscript Name} is defined as before.

If a tape is mounted on the tape drive and ready for recording, the system acts on the command by displaying the response "TAPE DUMP IN PROGRESS"; otherwise, an appropriate error message is displayed. The system begins writing the constituent files of the manuscript to tape, reordering feature descriptor numbers to form a continuous sequence in the process. When the operation is complete, the message "TAPE DUMP COMPLETED" is sent to the operator, who is then able to transport the tape to the Univac computer to be merged with terrain coordinate data.

DUMP MANUSCRIPT TO PRINTER

The operator can receive the hardcopy of a manuscript or partial manuscript by typing a command of the form

DUMPPrint {Manuscript Name}

If the printer is already in use, the message "PRINT REQUEST IN QUEUE" is returned. When the printer again becomes idle, the manuscript is given priority over all FADT files waiting in queue. Unlike the case of transferring manuscripts to tape, a number of manuscripts may be queued for printing and output on a first-in, first-out basis.

When the system has finished outputting the manuscript, a message of the form

PRINT OF {Manuscript Name} COMPLETE

is displayed.

5.2.6 Off-Line Functions

The only specific off-line function identified as necessary in the baseline system is maintaining the list of authorized users. To the list of off-line functions, we must add two: 1) general program development support, and 2) miscellaneous utility functions for manipulating files and controlling I/O devices. We are, of course, assuming a computer-based implementation. In this context, the above requirements translate into a requirement for operating system software.

We have been using the term off-line to refer to functions having low frequencies of use, i.e., much less than once a day. We wish to avoid giving the impression that off-line functions are exercised when no other activity is in process in the system--at night, for example. Rather, we intend that functions classified as off-line should take place concurrently--in a background mode, so to speak--with on-line functions. Off-line functions can be exercised only from the operator's position.

5.3 VOICE INPUT FUNCTIONS

This subsection describes the functions that are added to the baseline system of Subsection 5.2 to accommodate spoken inputs. The only ability we introduce into the design is that to exercise features of the baseline system through voice as well as on the keyboard. As discussed below, only a subset of the original features will be accessible by spoken command, and this alternate form of input shall be made available to the analyst only and not to the operator. These restrictions emphasize our conviction that, in this application, the proper role of voice input is as an alternative to, and back-up for, manual means, and that it is justified only if it will see a moderate-to-high frequency of use.

As envisioned, the analyst's station will be augmented by some form of speaker-dependent voice recognition device, which will be functionally independent of those at every other station; that is, groups of analysts will not need to compete for use of, for example, a common microphone, nor will the system be forced to discern the identity of analysts by comparing utterances. Thus, the speech recognition device, whatever its implementation, will appear to the analyst to be a privately held resource.

Data entry capabilities available through voice will be found to duplicate what the analyst can do with the keyboard. In addition, the analyst will be able to exercise all file review functions and will have access to informational aids. In all cases, he/she will employ a vocabulary of command and feature analysis keywords that is nearly identical in content and syntax to the typed inputs described in Subsection 5.2.

Session-control and mode-setting functions will continue to be available by keyboard alone. These functions are in an executive class, which means they involve the control of the analyst's interface with the system during the session. The consequences of a command being incorrectly interpreted are, therefore, potentially more serious when these functions are involved than for the category of functions encompassing data entry, file review, and informational aids. The susceptibility of voice recognition equipment to inadvertent inputs, caused by pickup of background noise, balance in favor of the more disciplined method of manual access for the former functions. Practically speaking, little loss of efficiency accrues as a result of this decision, considering that session-control and mode-setting functions are called on infrequently--typically only at the start and end of a session.

When the voice input capability is enabled, the analyst can make most command and data entries by uttering a single keyword. If multiple keywords are required, the analyst terminates his/her utterances with the word "ENTER". As a means of verification for the analyst, the system writes each keyword on the visual display as it is recognized. The system responds to an unrecognizable input by displaying some error message, such as "UNKNOWN", in place of the intended keyword. When the unrecognized word occurs at the beginning of the current command or data entry line, the system simply ignores it and prepares to accept new input without need for corrections. On the other hand, should the error occur after the current command or data entry has begun, the analyst is given the opportunity to correct the entry and proceed.

The analyst will be able to correct an entry within the current command or data line, for example, by speaking the word "CORRECTION". The effect of this command is analogous to the action of backspacing the cursor when using the keyboard entry: erasing the last utterance, whether a valid keyword or an "UNKNOWN", from the visual display and having it ignored by the system.

The analyst will be able to employ keyboard entry at all times, even if voice input has been enabled. To emphasize the overlap of voice and manaul input capabilities, both methods can be employed within the same command or data entry line. For instance, a multiple-keyword command begun by a spoken keyword may be completed by manually entering the remaining keyword or keywords from the keyboard, and vice versa. This feature is useful if certain spoken keywords become difficult to recognize consistently when the analyst has a cold or some other impairment of the vocal tract. In such a case, the analyst may elect to use keyboard entry for troublesome keywords, rather than retrain the spoken vocabulary or make repeated corrections.

A summary of on-line commands that pertain to voice input is presented in Table 9. The keyword vocabulary used in conjunction with these capabilities is the same set presented in the DLMS product specifications, supplemented by a small set specific to voice input such as those shown in Table 10.

5.3.1 Session Control/Mode Set

Although no session-control or mode-setting functions may be commanded by voice, activities associated with enabling/disabling voice input and vocabulary training are comprised by this category and are therefore controlled by the keyboard.

Vocabulary Training

A speaker-dependent voice recognition device works by correlating processed spoken inputs with an internally stored representation of each utterance in its vocabulary. These internal references are generated during the training process, or vocabulary training, which involves repeating each word in the vocabulary to the system several times.

The analyst enters the training mode by typing the command, "TRain" at the keyboard. The system responds by displaying the vocabulary index on the visual display. The vocabulary index is a tabular list that enumerates the various subvocabulary files stored in the system. Each subvocabulary file is represented in the vocabulary index by an identifying number and a brief phrase suggesting its content. The spoken vocabulary is subdivided into a number of subvocabularies because it is not possible, in general, to present the entire vocabulary at one time on the visual display.

TABLE 9. SUMMARY OF SAMPLE COMMANDS RELATING TO VOICE INPUT

CATEGORY	COMMAND	FORMAT
SESSION CONTROL AND MODE SETTING	* ENTER TRAIN MODE	TRain
	* REVIEW SUBVOCABULARY	VOcabulary {Subvocabulary Number}
	* PROMPT TRAINING	PROmpt {Word Number or Subvocabulary Number} ···
	* EXIT TRAIN MODE	Exit
	* SET REJECT THRESHOLD	REject {Level}
	* ENABLE VOICE INPUT	VIEnable
	* DISABLE VOICE INPUT	VIDisable
LINE AND FEATURE REVIEW	REVIEW FEATURE DESCRIPTOR	FEATURE {Digit} ··· ENTER
	REVIEW FIRST FEATURE DESCRIPTOR	FIRST
	REVIEW LAST FEATURE DESCRIPTOR	LAST
	REVIEW TOP LINE OF FEATURE	TOP
	REVIEW BOTTOM LINE OF FEATURE	BOTTOM
	REVIEW CURRENT LINE OF FEATURE	CURRENT
	REVIEW LINE ABOVE CURRENT LINE	UP
	REVIEW LINE BELOW CURRENT LINE	DOWN
ENTRY/ EDIT	ENTER INPUT MODE	INPUT
	EXIT INPUT MODE	ESCAPE
INFORMATIONAL AND COMPUTATIONAL AIDS	CONVERT CODE TO KEYWORD	CONVERT {Digit} ··· ENTER
	CONVERT KEYWORD TO CODE	CONVERT {Keyword} ENTER
	DELINEATE RANGE OF PARAMETER VALUES	RANGE ENTER
	DELINEATE FEATURE ID'S	RANGE {Digit} ENTER

*Entered via keyboard only.

TABLE 10. SAMPLE SPOKEN COMMAND KEYWORDS

ENTER	CONVERT
CORRECTION	RANGE
FEATURE	ZERO
FIRST	ONE
LAST	TWO
TOP	THREE
BOTTOM	FOUR
CURRENT	FIVE
UP	SIX
DOWN	SEVEN
INPUT	EIGHT
ESCAPE	NINE

The analyst can review a given subvocabulary by typing a command of the form

VOcabulary {Subvocabulary Number}

where {Subvocabulary Number} is the number that identifies the particular subvocabulary of interest. The system responds to this command by displaying the list of words (and/or phrases) that constitute the subvocabulary, along with a number identifying each entry. Within the entire vocabulary, numbers identifying subvocabularies and words are unique.

The analyst may prompt the system to train a single word or an entire subvocabulary. The training process is initiated when the analyst identifies the words to be trained by entering a command of the form

PROmpt {Word Number or Subvocabulary Number} ...

where {Word Number or Subvocabulary Number} is the number identifying the word or subvocabulary to be trained. Multiple words or subvocabularies may be specified by concatenating their number on the same line, separated by spaces.

Once the command is given, the system is ready to accept samples of each word (or phrase) to be trained. The system prompts the analyst by displaying each word on the visual display and waiting for the spoken response. Each word is called for several times until all words have been trained, at which time the system returns the message "TRAINING COMPLETE".

The analyst may continue to train additional words, exiting the train mode when he/she finishes by typing "EXIT". The system stores the amended file of vocabulary-training samples on-line for subsequent use.

Controlling Recognition Criteria

A speaker-dependent voice recognition system discriminates words by keeping some kind of score which judges how close a match is obtained between the current input and each of the reference samples in its vocabulary. It is not sufficient to assume that the word that matches best is the current input, since that word may only be the best of several poor matches, and any random input will correlate to some degree with one or more vocabulary words. A figure of merit, which we will call the reject level, is generally used as a criterion for judging that a recognition has occurred. The reject level is a numerical score that either represents an absolute level of matching or the difference between how well the current input correlates with the vocabulary word that matches best and the word that matches next best. When the reject level equals or exceeds a threshold value that can be set by the analyst, a recognition is declared.

The analyst sets the reject level threshold by entering a command at the keyboard of the form

REject {Level}

where {Level} is a number specifying the reject threshold. This value usually depends on the actual speech recognition equipment selected for use.

Enabling/Disabling Voice Input

When the analyst is prepared to begin entering commands and/or data by voice, he/she enables this capability by typing the command "VIEnable". The system responds by issuing the message "VOICE INPUT ENABLED".

If the analyst needs to disable the capability--for instance, when background noise rises to an unacceptable level or when he/she needs to converse with someone--he/she does so by typing "VIDisable". The system responds with the message "VOICE INPUT DISABLED".

5.3.2 Review Functions

All feature and line review functions discussed in Subsection 5.2.2 may be activated by means of voice command. The format of the various commands, Table 9, is virtually identical to the typed form of the commands in Table 7, and we will not elaborate on the use of each.

All but one command is effected by uttering a single keyword, and no terminating utterance is necessary. To review a feature descriptor at random, however, it is necessary to utter the keyword "FEATURE" followed by one to four digits specifying the feature number, followed by the terminating word "ENTER". Prior to terminating the command, the analyst can correct any of his/her inputs by using the word "CORRECTION".

5.3.3 Entry/Edit Functions

Entry and edit capabilities parallel those of the baseline system. As a comparison of Table 9 and 7 shows, entering and exiting the input mode is accomplished by the same keywords. Data entries may all be made as single utterances of either a number or an appropriate feature analysis keyword.

Some fields allow the option of entering multidigit codes in place of single keywords. While the latter is optimum from the standpoint of data entry speed and accuracy, the former will only require training of the ten digits rather than the roughly 260 feature ID's and other analysis keywords.

Thus, it is recommended that initial use of the voice input system be limited to numeric codes to ease the analyst's familiarization process with data entry by spoken utterances.

When a multidigit entry is made, the entry must be terminated with the keyword "ENTER". The word "ENTER" can be used alone if the analyst wishes to bypass a field of the feature descriptor without supplying an entry.

5.3.4 Other Functions

All of the informational aids available by keyboard are also available by voice command. Note from Table 9 that the RANGE command must be terminated in both its long and short forms. Computational support is not offered in voice, however, because arithmetic expressions can be represented much more concisely and unambiguously in written form.

The operator position has no speech input capabilities of its own. Furthermore, no on-line actions relating to analysts' speech recognition capabilities are required of the operator.

Alterations to the vocabulary index and to subvocabulary files shall be accomplished off-line, using the general-purpose file editing facilities of the operating system.

5.4 VOICE RESPONSE FUNCTIONS

We consider here functions associated with enhancing the baseline system of Subsection 5.2 to generate spoken responses in addition to visual outputs. All the original functions of the baseline design have been preserved in this projected design, and voice output will be used merely as a backup for the visual display. As before, this capability will be extended to the analyst only.

The interchangeable relationship of vocal and manual input methods does not prevail for aural and visual methods of response. The visual display provides a static response that can be read by the analyst at his/her convenience, while a spoken response must be heard and remembered. Thus, aural responses impose a requirement on the analyst's attention although they can be very useful if his/her eyes are otherwise engaged. We have therefore designed the system to produce voice outputs simultaneously when visual responses are presented.

As with voice input, voice output will not be used to supplement session-control and mode-setting functions. File review features and informational aids involve lengthy responses that are not suitable for aural presentation. Computational support is also offered at the keyboard only. We have chosen, therefore, to limit the use of spoken responses to the task of providing verifications and prompts to the analyst while he/she is engaged in entering feature data.

Each analyst's station is augmented by a voice output device such as an earphone or miniature loudspeaker that operates independently of every other station in place of, for example, a common loudspeaker. This capability requires only the two commands shown in Table 11, both of which are entered at the keyboard.

TABLE 11. SUMMARY OF SAMPLE COMMANDS RELATING TO VOICE RESPONSE

Category	Command	Format
SESSION CONTROL/ MODE SET	*ENABLE VOICE OUTPUT	VOEnable
	*DISABLE VOICE OUTPUT	VODisable

*Entered via keyboard only.

When the analyst wishes to enable the capability, he/she does so by typing the command "VOEnable" at the keyboard. The system responds to the command by returning the message "VOICE OUTPUT ENABLED" to the visual display.

Voice response is activated only when the analyst is engaged in entering data into an FADT file. The data entry process is initiated by using the input command, discussed in Subsections 5.2 and 5.3. In Input mode the system prompts the analyst aurally by telling him/her the name of the feature descriptor field that is being pointed to by the line pointer. If the current field is already filled, the system will say the contents of the field to the analyst, either as a string of digits that make up a numerical code or as a feature analysis keyword, whichever corresponds to the entry method being used. When the analyst completes an entry, the system repeats the entry as a means of verifying that it correctly understood the analyst. If the analyst makes an invalid entry, the system responds by saying a word such as "ERROR".

The vocabulary spoken by the system consists of the feature analysis keywords specified in the DLMS product specifications along with a small set of words and phrases such as those shown in Table 12.

TABLE 12. SAMPLE SPOKEN COMMAND KEYWORDS

FAC NUMBER	NUMBER OF PYLONS
FEATURE TYPE	ZERO
SURFACE MATERIAL	ONE
PREDOMINANT HEIGHT	TWO
STRUCTURES PER SQUARE MILE	THREE
PERCENT TREES	FOUR
PERCENT ROOF	FIVE
FEATURE IDENTIFICATION	SIX
DIRECTIVITY OR ORIENTATION	SEVEN
LENGTH OR DIAMETER	EIGHT
WIDTH	NINE
LEVEL	ERROR

If he/she wishes to disable the voice response capability, the analyst enters the command "VODisable", to which the system responds by displaying the message "VOICE OUTPUT DISABLED".

The process of modifying the spoken vocabulary is accomplished as an off-line activity.

6. OPERATIONAL SCENARIO

We define here a scenario for modeling an operational VDE system. This scenario is used in Section 7 to develop performance estimates by which to judge candidate implementations of the system design. This section also configures an initial system architecture, which will be refined in Section 7.

6.1 METHOD OF ANALYSIS

As constituted in Section 5, the VDE system operates almost entirely in response to the asynchronously entered commands of its user community, the analysts and operator. The system can be viewed as a queue of service requests--operator- and analyst-initiated commands--that arrive independently of one another (interarrival times can be characterized as a Poisson distribution) and are processed on a first-come, first-served basis.

The time taken by the system to service a request varies according to the request type. Each of the various commands available to the operator and analysts requires a time in process, which is somewhat predictable. The aim of this section is to develop a scenario that describes this distribution of commands, and from this distribution, to estimate the resulting distribution of service times.

The principal parameter of interest for the distribution of service requests is the average arrival rate λ. For the distribution of service times, the main parameters of interest are the average service time $\overline{X}$ and its second moment $\overline{X^2}$.

Given these parameters, we can calculate the average time taken by the system to service a request (that is, to recognize that a particular analyst has made a request of it) using the equation*

$$W = \frac{\lambda \overline{X^2}/2}{1-\rho} \quad ,$$

where ρ, the utilization factor, is the fraction of time the system is busy servicing requests. The utilization factor is given by*

$$\rho = \lambda \overline{X} \quad .$$

---

*L. Kleinrock, Queuing Systems, Volume II: Computer Applications, John Wiley & Sons, New York, 1976.

Our first approximation of an architecture to implement a multistation voice data entry system is shown in Figure 12 as a first-level partition of the system into subsystems. As the figure shows, we wish the system to accommodate up to 100 analyst stations and one operator station. There is nothing special about the figure 100, but it is an easy number to work with and clearly exceeds the 42-station minimum required for IFASS.

An integrated keyboard/CRT terminal satisfies the baseline requirements of keyboard entry and visual display for the analysts as well as for the operator. Such terminals are inexpensive, compact, and have a simple serial interface that can be operated over a wide range of information rates approaching 20K characters per second (a rate that greatly exceeds the reading and typing capabilities of the people using them).

The functional requirements of the baseline design system are implemented by means of a general-purpose computer that acts as a processing node to which the entire complement of terminals is connected. The computer subsystem contains a general-purpose computer, facilities for on-line and off-line storage, and a printer.

Our conviction that voice recognition and voice response capabilities are best used in this application as part of a hybrid manual/voice entry system motivated us to envision a system architecture that allows them to be added in a modular fashion--also allowing the system to be upgraded as advanced voice recognition and response capabilities become available. This requirement for modularity led us to incorporate these capabilities into separate and independent subsystems whose activities are controlled by the computer subsystem.

The voice recognition subsystem receives analysts' spoken inputs at a microphone (preferably an earpiece microphone such as Lear-Siegler's EarCom*) at each analyst's station. The subsystem contains storage for speaker reference data for every vocabulary for each analyst using the system. In practice, some of this storage may be allowed to physically reside in the computer subsystem when it is not in use. A communication path with the computer will facilitate the transfer of these data, along with providing a

*The EarCom combines microphone and earphone functions in a single unit which is specially molded to fit snugly in the analyst's ear, much as a hearing aid would.

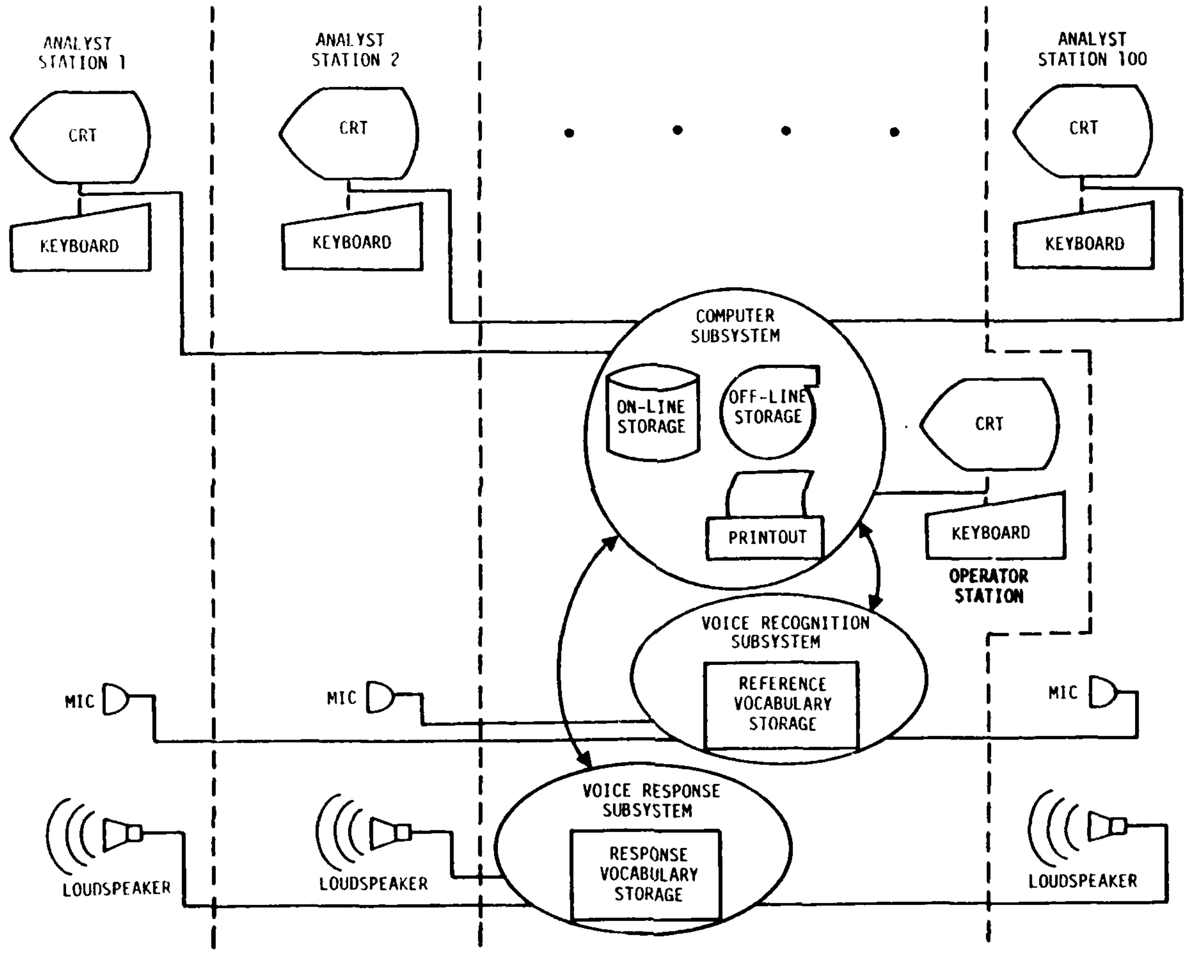


Figure 12. First-stage cut at system partition.

means for the voice recognition system to report data and receive control information.

In a similar manner, the voice response subsystem delivers its outputs by means of some type of earphone or loudspeaker separately installed at each analyst station. Storage is contained within the subsystem for data used in generating each word in the vocabulary of output utterances. This storage may also be allowed to physically reside in the computer subsystem when not needed, with a communication path to facilitate the transfer of storage from the computer subsystem and to deliver control inputs.

In this design, the computer subsystem becomes a centrally shared resource whose utilization will determine the performance characterisitics of the system. The aim of our analysis will be to guarantee sufficient processing capacity in the baseline system so that adequate reserve will remain even when advanced voice input and voice response capabilities are added.

6.2 A VDE SCENARIO

Figure 13 illustrates an operating scenario overview for a voice data entry system fully outfitted with both voice input and voice response capabilities. In the figure, the functions performed by the system are hierarchically decomposed and represented as a tree. Each node of the tree represents a major function, with the downward radiating branches showing the function's constituent subfunctions. The number enclosed by parentheses is the probability that the subfunction with which it is associated will be executed when the major function at the node is activated. These probabilities are, in effect, a measure of the duty cycle of the subfunctions of the major function. They are averages expected during typical operation of the VDE system.

From the figure, we can see that the scenario calls for low utilization of off-line as compared with on-line functions, by a ratio of nearly 1000 to 1. Furthermore, the same disparity exists between on-line functions exercised by the operator as opposed to those exercised by the group of analysts. From discussions with DMAAC personnel and from the results of the survey of DLMS analysts (reported in Section 2 and Appendix B), we can predict that no more than 20 percent of the analysts' time is spent on data entry. This is basically equivalent to saying that 20 percent or so of the possible 100 analysts will be actively using the system for data entry at any one time.

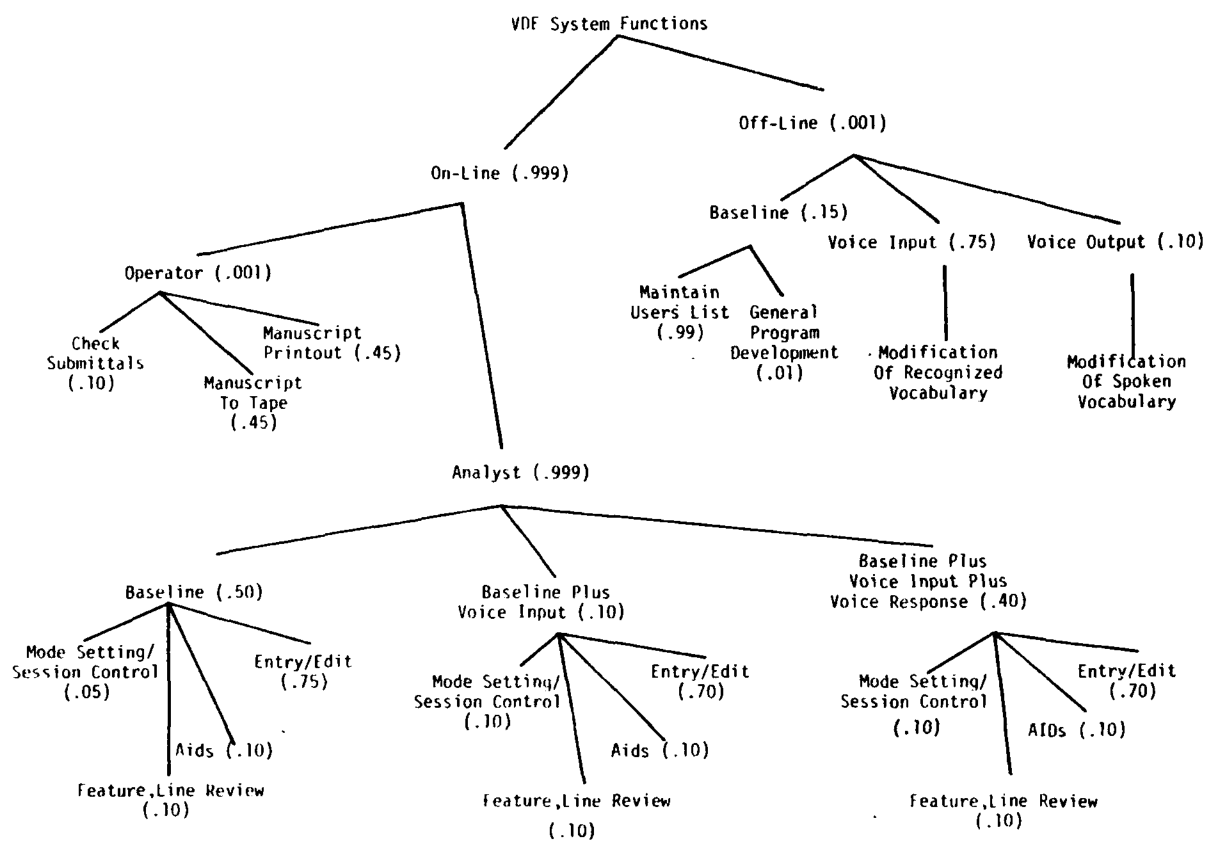


Figure 13. Functional hierarchy, showing utilization probabilities.

Within analyst functions, we assume that a significant group, 50 percent, will elect to use only the features of the baseline system. Another 40 percent will probably employ both voice input and voice response, while a minority, 10 percent will use the baseline plus just the voice input features. These percentages are estimates based on discussion with DMAAC personnel and on questionnaire responses. But actual percentages are not critical; what is important is the assumption that, in such a mixed-mode data entry system, different analysts will make use of manual or voice input/output options at different times, as their work requires.

Figures 14, 15 and 16 show the breakdown of analyst functions into their constituent commands for those three groups of analysts. Note that the principal differences between the groups are reflected in the percentage of commands that fall into the session-control and mode-setting category. When the extra capabilities of voice recognition and voice response are involved, the analyst spends slightly more of his/her time entering commands to control their use.

Table 13 enumerates all of the functions of the system, decomposed to their lowest level. Next to each is the cumulative, or absolute, probability of a function's being requested at any instant. Although these cumulative probabilities are specifically based on the scenario described in this section and the functional description that was developed in Section 5, none of these values has an overwhelming effect on system performance. Hence, a considerable variation in these parameters can exist without seriously affecting the performance measures discussed in Section 7.

6.3 AVERAGE REQUEST RATE

We can use the scenario we have developed, along with the statistics cited in Section 2, to derive the value of λ, the average request rate. It was estimated in Section 2 that at any given point in time, approximately 20 percent of the analysts are involved in the process of entering DLMS data. Given that our system is sized for 100 analysts, we can expect 20 users at a randomly chosen instant. It was also estimated that each analyst generates entries at a rate of one per minute on the average; so, the average arrival rate for service requests of the type corresponding to keyword or code entry will total 20 per minute for the entire system.

ANALYST FUNCTIONS
BASELINE SYSTEM
(.50)

- SESSION CONTROL/MODE SET (.05)
  - LOGIN (.15)
  - OPEN FADT FILE (.15)
  - CLOSE FADT FILE (.10)
  - SUBMIT FADT FILE (.05)
  - PRINT FADT FILE (.15)
  - DELETE FADT FILE (.05)
  - LIST FILE DIRECTORY (.10)
  - VERBOSE RESPONSE MODE (.07)
  - BRIEF RESPONSE MODE (.03)
  - LOGOUT (.15)
- ENTRY/EDIT (.75)
  - ENTER INPUT MODE (.10)
  - EXIT INPUT MODE (.10)
  - FEATURE CODE ENTRY (.60)
  - FEATURE KEYWORD ENTRY (.20)
- FEATURE,LINE REVIEW (.10)
  - REVIEW FEATURE DESCRIPTOR (.05)
  - FIRST FEATURE DESCRIPTOR (.05)
  - LAST FEATURE DESCRIPTOR (.15)
  - REVIEW TOP LINE OF FEATURE (.15)
  - BOTTOM LINE OF FEATURE (.15)
  - CURRENT LINE OF FEATURE (.15)
  - REVIEW LINE ABOVE CURRENT LINE (.15)
  - LINE BELOW CURRENT LINE (.15)
- AIDS (.10)
  - CONVERT CODE TO KEYWORD (.30)
  - CONVERT KEYWORD TO CODE (.05)
  - DELINEATE PARAMETER RANGE (.15)
  - DELINEATE FEATURE ID'S (.20)
  - EVALUATE EXPRESSION (.30)

Figure 14. Breakdown of analyst baseline functions.

ANALYST FUNCTION
BASELINE PLUS VOICE INPUT

(.10)

- SESSION CONTROL/MODE SET (.10)
  - LOGIN (.11)
  - OPEN FADT FILE (.11)
  - CLOSE FADT FILE (.07)
  - SUBMIT FADT FILE (.04)
  - PRINT FADT FILE (.11)
  - DELETE FADT FILE (.04)
  - LIST FILE DIRECTORY (.07)
  - VERBOSE RESPONSE MODE (.05)
  - BRIEF RESPONSE MODE (.02)
  - LOG OUT (.11)
  - ENTER TRAIN MODE (.01)
  - REVIEW SUBVOCABULARY (.03)
  - PROMPT TRAINING (.03)
  - EXIT TRAIN MODE (.01)
  - SET REJECT THRESHOLD (.01)
  - ENABLE VOICE INPUT (.12)
  - DISABLE VOICE INPUT (.06)
- ENTRY/EDIT (.70)
  - ENTER INPUT MODE (.10)
  - EXIT INPUT MODE (.10)
  - FEATURE CODE ENTRY (.40)
  - FEATURE KEYWORD ENTRY (.40)
  - AIDS (.10)
    - CONVERT CODE TO KEYWORD (.30)
    - CONVERT KEYWORD TO CODE (.05)
    - DELINEATE PARAMETER RANGE (.15)
    - DELINEATE FEATURE ID'S (.20)
    - EVALUATE EXPRESSIONS (.30)
- FEATURE,LINE REVIEW (.10)
  - REVIEW FEATURE DESCRIPTOR (.05)
  - FIRST FEATURE DESCRIPTOR (.05)
  - LAST FEATURE DESCRIPTOR (.15)
  - REVIEW TOP LINE OF FEATURE (.15)
  - BOTTOM LINE OF FEATURE (.15)
  - CURRENT LINE OF FEATURE (.15)
  - REVIEW LINE ABOVE CURRENT LINE (.15)
  - LINE BELOW CURRENT LINE (.15)

Figure 15. Breakdown of analyst functions, baseline plus voice input.

ANALYST FUNCTIONS
BASELINE PLUS VOICE INPUT PLUS VOICE RESPONSE
(.40)

SESSION CONTROL/MODE SET (.10)

- LOG IN (.09)
- OPEN FADT FILE (.09)
- CLOSE FADT FILE (.06)
- SUBMIT FADT FILE (.03)
- PRINT FADT FILE (.09)
- DELETE FADT FILE (.03)
- LIST FILE DIRECTORY (.05)
- VERBOSE RESPONSE MODE (.04)
- BRIEF RESPONSE MODE (.02)
- LOGOUT (.09)
- ENTER TRAIN MODE (.01)
- REVIEW SUBVOCABULARY (.03)
- PROMPT TRAINING (.03)
- EXIT TRAIN MODE (.01)
- SET REJECT THRESHOLD (.01)
- ENABLE VOICE INPUT (.09)
- DISABLE VOICE INPUT (.05)
- ENABLE VOICE OUTPUT (.09)
- DISABLE VOICE OUTPUT (.09)

ENTRY/EDIT (.70)

- ENTER INPUT MODE (.10)
- EXIT INPUT MODE (.10)
- FEATURE CODE ENTRY (.40)
- FEATURE KEYWORD ENTRY (.40)

AIDS (.10)

- CONVERT CODE TO KEYWORD (.30)
- CONVERT KEYWORD TO CODE (.05)
- DELINEATE PARAMETER RANGE (.15)
- DELINEATE FEATURE ID'S (.20)
- EVALUATE EXPRESSIONS (.30)

FEATURE,LINE REVIEW (.10)

- REVIEW FEATURE DESCRIPTOR (.05)
- FIRST FEATURE DESCRIPTOR (.05)
- LAST FEATURE DESCRIPTOR (.15)
- REVIEW TOP LINE OF FEATURE (.15)
- BOTTOM LINE OF FEATURE (.15)
- CURRENT LINE OF FEATURE (.15)
- REVIEW LINE ABOVE CURRENT LINE (.15)
- LINE BELOW CURRENT LINE (.15)

Figure 16. Breakdown of analyst functions, baseline plus voice input plus voice response.

TABLE 13. VDE FUNCTIONS AND THEIR INSTANTANEOUS PROBABILITY OF BEING REQUIRED (page 1 of 4)

FUNCTION	CUMULATIVE PROBABILITY
OFF-LINE	
MAINTAIN USERS LISTS	148.5×10^{-6}
GENERAL PROGRAM DEVELOPMENT	1.5×10^{-6}
MODIFICATION OF RECOGNIZED VOCABULARY	750.0×10^{-6}
MODIFICATION OF SPOKEN VOCABULARY	100.0×10^{-6}
ON-LINE, OPERATOR	
CHECK SUBMITTALS	99.9×10^{-6}
MANUSCRIPT TO TAPE	449.6×10^{-6}
MANUSCRIPT PRINTOUT	449.6×10^{-6}
ON-LINE, ANALYST, BASELINE	
SESSION CONTROL/MODE SET:	
LOGIN	3.74×10^{-3}
OPEN FADT FILE	3.74×10^{-3}
CLOSE FADT FILE	2.50×10^{-3}
SUBMIT FADT FILE	1.25×10^{-3}
PRINT FADT FILE	3.74×10^{-3}
DELETE FADT FILE	1.25×10^{-3}
LIST FILE DIRECTORY	2.50×10^{-3}
VERBOSE RESPONSE MODE	1.75×10^{-3}
BRIEF RESPONSE MODE	0.75×10^{-3}
LOGOUT	3.74×10^{-3}
FEATURE, LINE REVIEW:	
REVIEW FEATURE DESCRIPTOR	2.50×10^{-3}
FIRST FEATURE DESCRIPTOR	2.50×10^{-3}
LAST FEATURE DESCRIPTOR	7.49×10^{-3}
REVIEW TOP LINE OF FEATURE	7.49×10^{-3}
BOTTOM LINE OF FEATURE	7.49×10^{-3}
CURRENT LINE OF FEATURE	7.49×10^{-3}
REVIEW LINE ABOVE CURRENT LINE	7.49×10^{-3}
LINE BELOW CURRENT LINE	7.49×10^{-3}
ENTRY/EDIT:	
ENTER INPUT MODE	37.43×10^{-3}
EXIT INPUT MODE	37.43×10^{-3}
FEATURE CODE ENTRY	224.55×10^{-3}
FEATURE KEYWORD ENTRY	74.85×10^{-3}

TABLE 13--Continued (page 2 of 4)

FUNCTION	CUMULATIVE PROBABILITY
AIDS:	
CONVERT CODE TO KEYWORD	14.97×10^{-3}
CONVERT KEYWORD TO CODE	2.50×10^{-3}
DELINEATE PARAMETER RANGE	7.49×10^{-3}
DELINEATE FEATURE ID'S	9.98×10^{-3}
EVALUATE EXPRESSIONS	14.97×10^{-3}
ON-LINE, ANALYST, BASELINE PLUS VOICE INPUT	
SESSION CONTROL/MODE SET:	
LOGIN	1.10×10^{-3}
OPEN FADT FILE	1.10×10^{-3}
CLOSE FADT FILE	0.70×10^{-3}
SUBMIT FADT FILE	0.40×10^{-3}
PRINT FADT FILE	1.10×10^{-3}
DELETE FADT FILE	0.40×10^{-3}
LIST FILE DIRECTORY	0.70×10^{-3}
VERBOSE RESPONSE MODE	0.50×10^{-3}
BRIEF RESPONSE MODE	0.20×10^{-3}
LOGOUT	1.10×10^{-3}
ENTER TRAIN MODE	0.10×10^{-3}
REVIEW SUBVOCABULARY	0.30×10^{-3}
PROMPT TRAINING	0.30×10^{-3}
EXIT TRAIN MODE	0.10×10^{-3}
SET REJECT THRESHOLD	0.10×10^{-3}
ENABLE VOICE INPUT	1.20×10^{-3}
DISABLE VOICE INPUT	0.60×10^{-3}
FEATURE, LINE REVIEW:	
REVIEW FEATURE DESCRIPTOR	0.50×10^{-3}
FIRST FEATURE DESCRIPTOR	0.50×10^{-3}
LAST FEATURE DESCRIPTOR	1.50×10^{-3}
REVIEW TOP LINE OF FEATURE	1.50×10^{-3}
BOTTOM LINE OF FEATURE	1.50×10^{-3}
CURRENT LINE OF FEATURE	1.50×10^{-3}
REVIEW LINE ABOVE CURRENT LINE	1.50×10^{-3}
LINE BELOW CURRENT LINE	1.50×10^{-3}
ENTRY/EDIT:	
ENTER INPUT MODE	6.99×10^{-3}
EXIT INPUT MODE	6.99×10^{-3}
FEATURE CODE ENTRY	27.94×10^{-3}
FEATURE KEYWORD ENTRY	27.94×10^{-3}

AD-A100 470 TECHNOLOGY SERVICE CORP SANTA MONICA CA F/G 5/8
MULTISTATION VOICE DATA ENTRY CONFIGURATION STUDY.(U)
APR 81 P W GREGORY, J M REAVES F30602-80-C-0033
UNCLASSIFIED TSC-PD-B662-1 RADC-TR-81-50 NL

2 OF 2
AD A100470

END
DATE FILMED
7-81
DTIC

TABLE 13--Continued (page 3 or 4)

FUNCTION	CUMULATIVE PROBABILITY
AIDS.	
CONVERT CODE TO KEYWORD	2.99×10^{-3}
CONVERT KEYWORD TO CODE	0.50×10^{-3}
DELINEATE PARAMETER RANGE	1.50×10^{-3}
DELINEATE FEATURE ID'S	2.00×10^{-3}
EVALUATE EXPRESSION	2.99×10^{-3}
<u>ON-LINE, ANALYST, BASELINE PLUS VOICE INPUT PLUS VOICE RESPONSE</u>	
SESSION CONTROL/MODE SET:	
LOGIN	3.59×10^{-3}
OPEN FADT FILE	3.59×10^{-3}
CLOSE FADT FILE	2.40×10^{-3}
SUBMIT FADT FILE	1.20×10^{-3}
PRINT FADT FILE	3.59×10^{-3}
DELETE FADT FILE	1.20×10^{-3}
LIST FILE DIRECTORY	2.00×10^{-3}
VERBOSE RESPONSE MODE	1.60×10^{-3}
BRIEF RESPONSE MODE	0.80×10^{-3}
LOGOUT	3.59×10^{-3}
ENTER TRAIN MODE	0.40×10^{-3}
REVIEW SUBVOCABULARY	1.20×10^{-3}
PROMPT TRAINING	1.20×10^{-3}
EXIT TRAIN MODE	0.40×10^{-3}
SET REJECT THRESHOLD	0.40×10^{-3}
ENABLE VOICE INPUT	3.59×10^{-3}
DISABLE VOICE INPUT	2.00×10^{-3}
ENABLE VOICE OUTPUT	3.59×10^{-3}
DISABLE VOICE OUTPUT	3.59×10^{-3}
FEATURE, LINE REVIEW:	
REVIEW FEATURE DESCRIPTOR	2.00×10^{-3}
FIRST FEATURE DESCRIPTOR	2.00×10^{-3}
LAST FEATURE DESCRIPTOR	5.99×10^{-3}
REVIEW TOP LINE OF FEATURE	5.99×10^{-3}
BOTTOM LINE OF FEATURE	5.99×10^{-3}
CURRENT LINE OF FEATURE	5.99×10^{-3}
REVIEW LINE ABOVE CURRENT LINE	5.99×10^{-3}
LINE BELOW CURRENT LINE	5.99×10^{-3}

TABLE 13--Concluded (page 4 of 4)

FUNCTION	CUMULATIVE PROBABILITY
ENTRY/EDIT:	
ENTER INPUT MODE	27.94×10^{-3}
EXIT INPUT MODE	27.94×10^{-3}
FEATURE CODE ENTRY	111.78×10^{-3}
FEATURE KEYWORD ENTRY	111.78×10^{-3}
AIDS:	
CONVERT CODE TO KEYWORD	11.98×10^{-3}
CONVERT KEYWORD TO CODE	2.00×10^{-3}
DELINEATE PARAMETER RANGE	5.99×10^{-3}
DELINEATE FEATURE ID'S	7.98×10^{-3}
EVALUATE EXPRESSION	11.98×10^{-3}

From Table 13 it can be calculated that the functions involving data entry (feature keyword entry, feature code entry) have an aggregate probability of utilization of 578.84×10^{-3}. Since the probability of use of a function is expressed as a ratio of the number of times it is requested to the total number of functions requested over an interval of time, we can determine λ as follows:

$$\lambda = \frac{\text{Average Arrival Rate of Entry Function Requests}}{\text{Probability of Arrival of Entry Function Requests}}$$

$$= \frac{20/\text{minute}}{578.84 \times 10^{-3}} = 34.55 \text{ Requests per Minute}$$

$$= 575.9 \times 10^{-3} \text{ Requests per Second}$$

During periods of peak activity, this average request rate can be expected to at least double (see Section 2). In addition, if it happened that all 100 analysts were entering data at the same instant, then the peak request rate could increase to ten times the average rate. Of course, actual experience with the IFASS system could provide far more realistic on-line data entry rates. Thus it is recommended in Section 8 that provisions be made in the IFASS software to monitor statistics on data entry/response rates.

7. POTENTIAL CONFIGURATIONS

In this section, we shall present three different approaches to implementing the VDE system. The first is a centralized configuration employing a single central minicomputer in the computer subsystem. The second configuration employs multiple minicomputers to distribute the available processing capacity and add to reliability. The final configuration utilizes distributed processing at its utmost with completely independent stations, each having its own microcomputer and voice I/O devices.

We shall describe each configuration, develop performance estimates using the model of Section 6, and evaluate them against each other.

7.1 CENTRALIZED CONFIGURATION

Figure 17 illustrates the first configuration for implementing a multistation voice data entry system. Relative to the partitioned system presented in Section 6 (Figure 12), this configuration is characterized by a processing subsystem containing a single central minicomputer and by voice recognition and voice response subsystems whose components are dispersed among the analyst's stations.

Computer Subsystem

The computer subsystem is highlighted by a high-performance, general-purpose minicomputer that acts as both data processor and controller for the rest of the system. The computer's resources include peripherals for on-line and off-line storage, hardcopy printout, and communication interfaces.

The CPU's word length is 16 bits, and it is capable of performing both single and double precision arithmetic on integer and floating point data. No special-purpose floating point hardware is required because of the relatively low rate of utilization of the system's computational support capabilities. The CPU receives notification of the activities of its peripherals by means of a hardware-vectored interrupt network. This feature relieves the CPU of the burden of constantly polling its peripherals to detect activity. Direct memory access (DMA) capability is also provided to allow high-speed data transfers to occur between selected peripherals and main memory without requiring constant attention by the CPU.

The CPU has a system console consisting of a keyboard/CRT terminal. The system console acts as a privileged terminal for use in controlling the

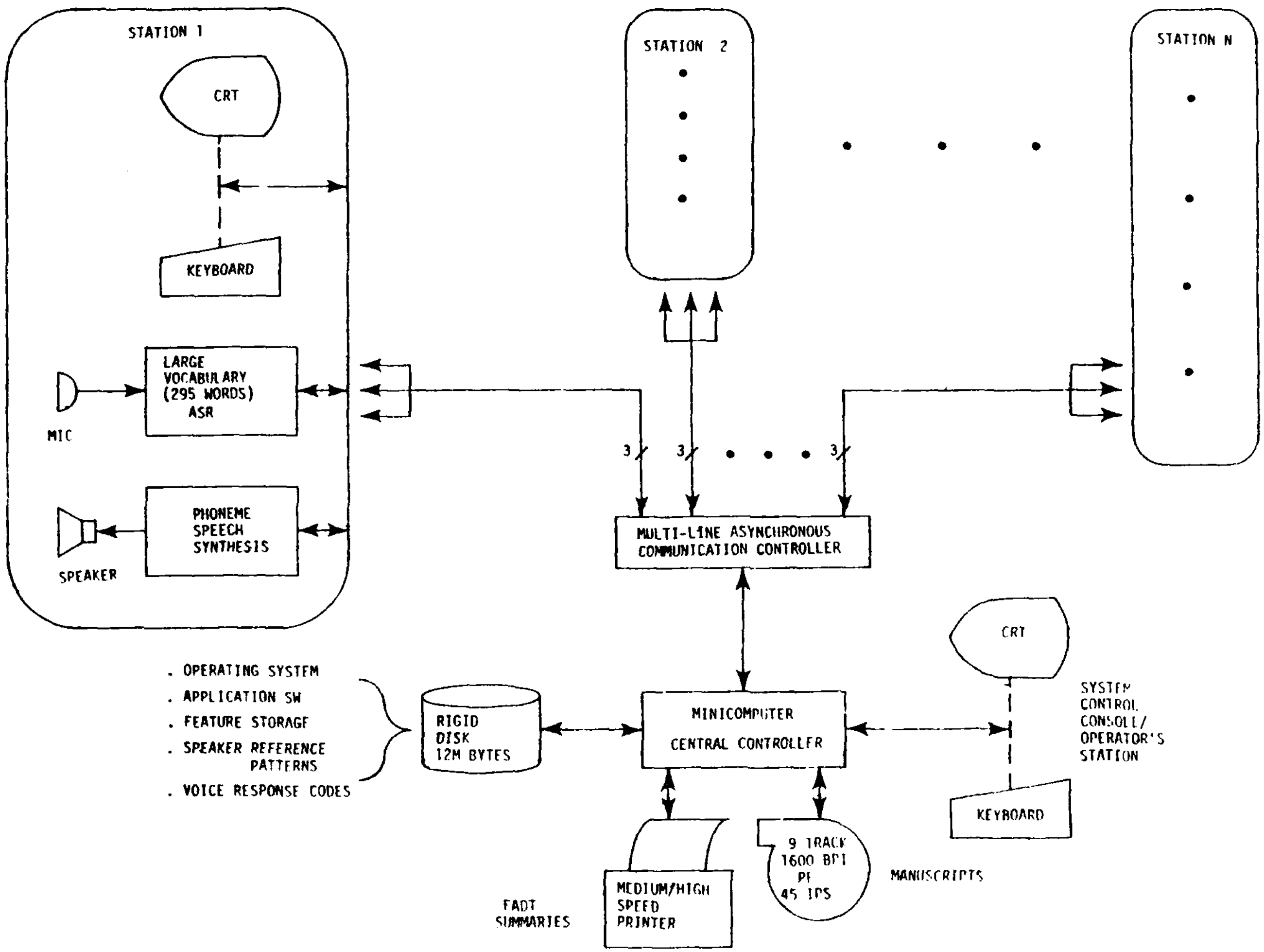


Figure 17. DLMS voice data entry system: centralized approach.

operation of the computer subsystem and for running diagnostic checks. It has general-purpose uses in addition, and we propose to use it as the operator's station. It is connected via an asynchronous serial RS-232 interface.

On-line storage capabilities are provided by randomly accessible main memory within the processor and by an external disk drive unit. Main memory is composed of MOS semiconductor devices as opposed to magnetic cores for high density and short access times. (The volatility of semiconductor storage is offset by the availability of nonvolatile storage on the disk unit.) The disk drive is a rigid, moving head design containing multiple platters, and is interfaced to the CPU by means of a disk controller that is integral to the processor chassis. The disk controller uses one of the high bandwidth data channels provided by the DMA capability, cited earlier. Flexible discs were not selected for this configuration because of their low storage capacity and slow transfer rates as compared with rigid disc drives. Fixed head disks allow faster data access, but were not chosen, owing to their greater cost and complexity.

A tape transport serves as the primary means of off-line storage. The transport chosen is a 9-track unit using phase encoding of data at a density of 1600 characters (bytes) per inch and operating at a speed of 45 inches per second. Tape will be utilized as the medium for transporting completed manuscripts from the system to the Univac computer for final processing. The characteristics of the transport were chosen for compatibility with the format used by the Univac system. The transport interfaces to the CPU by means of a tape controller, and like the disc controller, a DMA channel is employed to facilitate data transfers.

Secondary off-line storage capability is effected by using the removable platters of the disk drive unit. This is available for storage of applications software as may be required for general program development activities, or in case of failure of the tape drive, for storage of FADT manuscripts. In the latter case, manuscripts could be saved for later transfer to tape when the tape unit is repaired, or they could be transferred [illegible] to the Univac, if a compatible drive exists there.

The printer could be either a medium-speed [illegible] high speed line printer, as desired. A charac[illegible] at 120 characters per second can print an [illegible] [illegible]taining 1500 features in their coded form (without English headings for descriptor fields, and without keywords) in seven minutes or less. For more elaborate printouts, a high-speed printer would be justified. The medium-speed printer uses an asynchronous serial RS-232 interface and is, therefore, simpler to install than the high-speed printer that requires a parallel interface and a printer controller.

The computer subsystem connects with the remainder of the system via asynchronous, serial RS-232 interfaces effected by means of a multichannel asynchronous communications controller. A set of three such interfaces is used at each analyst's station, one connecting to the keyboard/CRT terminal, one to a speech recognition terminal, and one to a voice response terminal. The communication controller is a modularly expandable device that can be programmed to operate at a wide variety of data rates from 10 to 1920 characters per second. Interfaces can be added incrementally using this approach. For instance, a [illegible] line system could be configured initially with keyboard/CRT terminals [illegible] voice recognition and voice response equipment could be added lat[illegible]

Voice Recognition Subsystem

The voice recognition terminals operate in [illegible]. They will each provide their own on-line [illegible] the entire vocabulary of analyst utterances [illegible] however, the computer subsystem will provide [illegible]ference, uploading them when voice recognition [illegible] subsequent to LOGON, and downloading them when [illegible] result of training. The advantage of storing [illegible] any analyst can then work at any station, if so desired. [illegible]tate uploading the large amount of data corresponding to the [illegible] vocabulary, the interface would be programmed to operate at the [illegible] rate of 1920 characters per second. Thus, uploading the present vocabulary of approximately 300 keywords could be accomplished in under 20 seconds, assuming that no more than 130 characters are required to represent each one (as is typical of the devices we have surveyed). During operation, recognitions are reported by sending coded messages to the computer subsystem. A typical message might involve 5 characters or less, requiring very light utilization of the interface.

In order to store the entire vocabulary at each analyst's terminal, a large-capacity voice recognition unit must be used. In practice, such units are themselves built around a minicomputer, and typically service four users (depending on the total vocabulary size). However, a voice recognition unit with a smaller on-line vocabulary could, in the [illegible] the application. This would require uploading [illegible] vocabulary at appropriate intervals, subject to [illegible] undertaken by the analyst. The trade-off [illegible] cost versus higher loading of the central computer [illegible] response times.

Software and [illegible]

[illegible] necessary to operate the system consists of an operating [illegible] application package. The operating system contains, at a [illegible] the following components: 1) editor; 2) high-level language compiler; 3) assembler; 4) linker-loader; 5) diagnostics; 6) I/O utilities; and 7) executive. Candidate high-level languages include FORTRAN, favored because of its widespread use and Department of Defense endorsement, and PASCAL, a language rapidly gaining acceptance for its adherence to structured design. A real-time, multitasking executive program keeps track of the asynchronously occurring events taking place both within and external to the computer system.

Table 14 enumerates the requirements for on-line storage within the computer subsystem. To allow for at least 25 percent excess capacity in main memory and disk storage, a total of 400K bytes of main memory and 12M bytes of disc are required.

Performance

In Table 15, a summary of the major functions of Table 13, and the estimated number of assembly-language computer instructions required to execute each, is presented. These estimates are based on the types of instructions that characterize a broad class of minicomputers and microcomputers.

The focus of our performance analysis is on the computer subsystem, as it is the only shared resource in the system. From Section 6, the average arrival rate of service requests, λ, was found to be 575.9×10^{-3} operations or functions per second (i.e., about 34 requests per minute). From Table 15, the average contribution of each function to the total system loading is obtained by multiplying its cumulative probability times its instruction count. The average service time, $\overline{X}$, is found by multiplying the summation of these contributions times the average number of instructions per second that can be executed by the computer subsystem. A figure of 500K instructions per second, typical for minicomputers, yields an average service time of 23.4 msec per operation.

TABLE 14. ESTIMATED ON-LINE STORAGE REQUIREMENTS, CENTRALIZED CONFIGURATION

Category	Basis	Main Memory	Disc Memory
OPERATING SYSTEM	Source + Object + Load Module		3M bytes
	Load Module + Buffers	120K bytes	
APPLICATION SW	Source + Object + Load Module		500K bytes
	Load Module + Buffers	200K bytes	
SPEAKER REFERENCE PATTERNS	300 words x 128 bytes x 100 analysts		3.85M bytes
FADT FILES	2 files x 100 features x 30 characters x 100 analysts		600K bytes
MANUSCRIPTS	10 x 1500 features x 30 characters		450K bytes
TOTALS		320K bytes	8.4M bytes

TABLE 15. FUNCTION PROBABILITY AND PROCESSING REQUIREMENTS, CENTRALIZED CONFIGURATION

Function	Average Processing Requirements (Instructions)
OFF-LINE	15.4M
ON-LINE, OPERATOR	2.5M
ON-LINE, ANALYST, BASELINE:	
Session Control/Mode Set	300K
Feature, Line Review	46K
Entry/Edit	14K
AIDs	45K
ON-LINE, ANALYST, BASELINE PLUS VOICE INPUT:	
Session Control/Mode Set	940K
Feature, Line Review	41K
Entry/Edit	12K
AIDs	43K
ON-LINE, ANALYST, BASELINE PLUS VOICE INPUT PLUS VOICE RESPONSE:	
Session Control/Mode Set	1.0M
Feature, Line Review	41K
Entry/Edit	24K
AIDs	43K

The utilization factor, ρ, for the computer subsystem is then given by

$$\rho = \lambda\overline{X} = (575.9 \times 10^{-3})(23.4 \times 10^{-3}) = 0.014 = 1.4\% \quad .$$

Thus, under the scenario developed in Section 6, 98.6 percent excess capacity remains in the computer subsystem during average use. Under worst-case, peak-period usage (all analysts entering data at peak rates), the utilization factor could increase by an order of magnitude, which still leaves 86 percent excess capacity.

The average waiting time is found from the second moment of service time, $\overline{X^2}$, and is given by

$$W = \frac{\lambda\overline{X^2}/2}{1 - \rho} = 3.5 \text{ msec} \quad .$$

Therefore, the system can be expected to give virtually instantaneous service to each incoming request. Under worst-case, peak-period usage, this average waiting time would increase to about 40 msec, which is still instantaneous as far as the user is concerned.

Cost

The hardware cost breakdown for the centralized configuration is given in Table 16. Cost estimates are typical figures for the type of equipment specified and do not include any system engineering, installment, maintenance, or software development costs. Total hardware cost is approximately $15K per analyst for a full-capacity system containing 100 stations. The greatest contributor to cost is the voice recognition hardware, and if limited-capacity voice recognition equipment was used (as mentioned previously) total hardware costs could easily be halved. The baseline system itself could be configured for as little as $2K per analyst in hardware costs.

7.2 CONFIGURATION EMPLOYING TWO-LEVEL CENTRALIZATION

Figure 18 illustrates an alternative configuration for the voice data entry system, one that employs a two-level approach to the computer subsystem. This configuration has greater reliability than the centralized configuration, and is characterized by multiple computers in the computer subsystem and a small-vocabulary speech recognition device.

TABLE 16. TYPICAL HARDWARE COSTS, CENTRALIZED CONFIGURATION

Item	Basis	Cost
COMPUTER SUBSYSTEM	CPU + Peripherals	$ 75K
ANALYST STATIONS (BASELINE)	$1K x 100	$ 100K
OPERATOR STATION	$1K x 1	$ 1K
VOICE RECOGNITION SUBSYSTEM	$10K x 100	$1,000K
VOICE RESPONSE SUBSYSTEM	$3K x 100	$ 300K
TOTAL		$1,476K

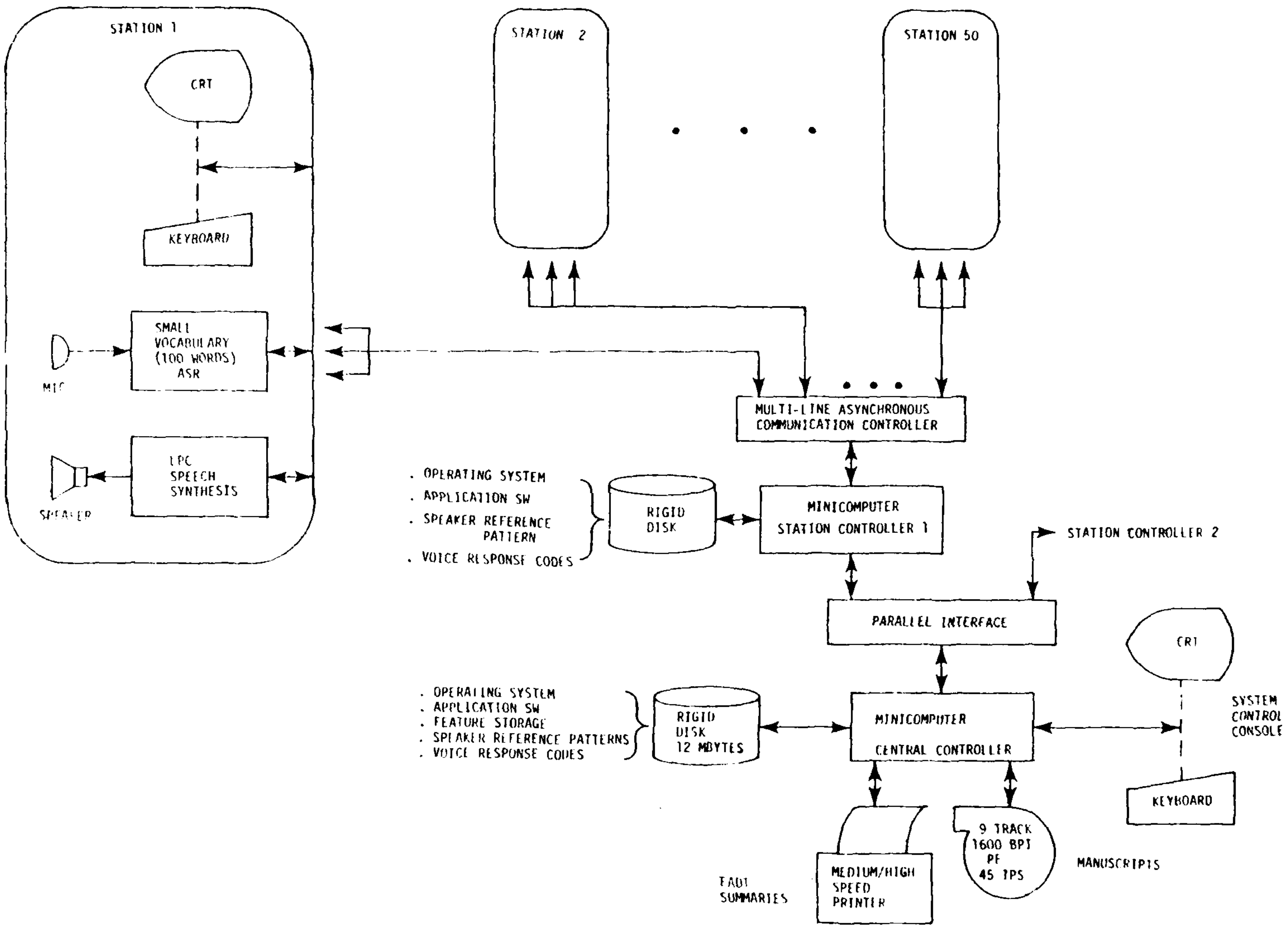


Figure 18. DLMS voice data entry system: two-level centralization approach.

Computer Subsystem

The computer subsystem contains three minicomputers for greater reliability than that provided in the centralized configuration of Subsection 7.1. That is, when the central minicomputer goes down, data entry essentially halts for all analysts in the centralized configuration, but is unaffected in this two-level configuration (though other functions would be affected). Greater processing capacity is also available in this configuration.

In the two-level approach, two of the three minicomputers operate as station controllers (each tending half of the system's 100 stations). The station controllers provide most analyst services autonomously. Facilities are maintained to provide temporary storage for open FADT files and reference patterns making up the recognition vocabularies of its analysts. The response vocabulary and the list of authorized users are stored redundantly in each station controller.

The central controller is essentially the same as that described in Subsection 7.1. It supports the operator's station and provides archival storage for the system's central data base. Portions of the data base--authorized users list, voice response codes--are uploaded to the station controllers upon initializing the system and when they are modified by the operator. Voice recognition reference samples are uploaded when the analyst enables voice response for the first time during a session; and after training, the updated samples are downloaded back to archival store.

The on-line functions performed by the central controller are those directly supporting the operator and those relating to the use of the printer and tape transport. FADT files are downloaded from the station controllers when a printout is desired or when they are submitted for inclusion in a manuscript. FADT files are also stored permanently there when they are not being modified.

The use of three computers within the processing subsystem enhances the reliability of the system as a whole. If a station controller fails, then the system controller will temporarily take over its data entry functions; if the central controller itself fails, then manuscripts cannot be compiled nor printouts obtained, but all other activities can take place as usual, using just the station controllers while the central controller is being repaired. No longer is the system vulnerable to a single point failure that would render it useless, as in the centralized configuration of Subsection 7.1.

The characteristics of the central and station controllers are essentially the same as those described in Subsection 7.1, with the exception of the differences in their complement of peripheral devices. The interface between the central and station controllers is a high-speed parallel interface utilizing direct memory access capability.

Voice Recognition Subsystem

The voice recognition subsystem utilizes a microcomputer-based voice recognition device that has a vocabulary of no more than 100 words rather than a more expensive minicomputer-based unit of several times that capacity. This can be accomplished in two different ways.

First, if the numeric feature ID codes are used exclusively in place of feature ID descriptions (suspension bridge, radar antenna, etc.), then the total vocabulary size drops from over 300 words to no more than 50 or so. (No more than 20 words would be required if numeric codes are used exclusively for all feature parameters.) However, one of the chief advantages of voice data entry is that descriptions can be used instead of numeric codes for faster, more accurate data entry. Hence, this option is somewhat counterproductive.

The second option requires that feature IDs be entered in a two-step process. The first step involves entering the feature's general category type (one of nine types, as indicated by the first digit of the 3-digit feature ID code), such as "Industry," "Transportation," or "Residential/Agricultural," for example. The second step involves specifying the particular feature within the general category just specified. For example, in the "Industry" category, the specific feature could be an offshore platform, a power plant, a strip mine, on any one of the other fifty features defined in this category.

At present, the "Industry" category has the largest subset of features (53), while the "Residential/Agricultural" category has the smallest subset, with only 11 features defined. The average number of features per category is currently 28, and while this number is likely to increase, a limit of 100 is sure to meet DLMS needs in the foreseeable future.

In practice, this two-step process would operate as follows. When the general category command is recognized by the system, the reference samples for the subset of features in that category are uploaded to the speech recognition terminal from the station controller. (If the analyst in unsure of the feature descriptions possible in this subset, he/she can check on the CRT screen, where the choices will be displayed.) After the specific feature description has been entered and verified, the feature subset remains at the analyst's station until a different general category type is specified, at which time the new feature subset is uploaded from the station controller, as before.

Software and On-Line Storage

The software required in this configuration is approximately the same as used in the centralized configuration of Subsection 7.1, with the exception that the application package is split into that portion that runs in the central controller and that which runs in the station controllers.

Table 17 details the on-line storage requirement for the central controller and station controllers. For at least 25 percent reserve capacity, the central controller requires approximately 200K bytes of main memory and 12M bytes disk storage. The station controllers require approximately 250K bytes of main memory and 7.5M bytes of disk storage.

Performance

Table 18 summarizes the major system functions and assigns instruction counts as they apply to the central controller and station controllers.

The station controllers exhibit an average service time, $\overline{X}$, of 36.8 msec. The utilization factor, ρ, for the station controllers is now:

$$\rho = \lambda\overline{X} = (575.9 \times 10^{-3})(36.8 \text{ msec}) = 0.021 = 2.1\% \quad .$$

Each station controller, then, has 97.9 percent excess processing capacity remaining during average usage. For worst-case, peak-period use, the utilization factor could increase to 21 percent, leaving 79 percent excess processing capacity. The average waiting time is found to be 3.0 msec under average conditions, and just under 40 msec for worst-case, peak-period usage.

The central controller exhibits an average service time of about 9 msec, from which its utilization is found to be 0.005. Therefore, its excess capacity is 99.5 percent. The corresponding waiting time is 1.4 msec.

TABLE 17. ESTIMATED ON-LINE STORAGE REQUIREMENTS, TWO-LEVEL CENTRALIZED CONFIGURATION

[illegible]ORY	BASIS	CENTRAL CONTROLLER		STATION CONTROLLER	
		Main Memory	Disk	Main Memory	Disk
OPE[illegible] [illegible]STEM	Source + Object + Load Module		3M bytes		3M bytes
	Load Module + Buffers	120K bytes			
APPLICA[illegible] [illegible]WARE	Source + Object + Load Module		500K bytes		
	Load Module + Buffers	50K bytes		200K bytes	40K bytes
SPEAKER REF[illegible] PATTERNS	300 words x 128 bytes x 100 analysts		3.85M bytes		
	300 words x 128 bytes x 50 analysts				1.9M bytes
FADT FILES	[illegible] files x 100 features x [illegible] characters x 100 analysts		600K bytes		
MANUSCRIPTS	[illegible]500 features x [illegible]aracters		450K bytes		
TOTALS		170K bytes	8.4M bytes	200K bytes	4.94M bytes

TABLE 18. FUNCTION PROBABILITY AND PROCESSING REQUIREMENTS, TWO-LEVEL CENTRALIZATION

FUNCTION	AVERAGE PROCESSING REQUIREMENTS (Instructions)	
	Central Controller	Station Controller
OFF-LINE	15.4M	27[illegible]
ON-LINE, OPERATOR	2.5M	
ON-LINE, ANALYST, BASELINE:		
Session Control/Mode Set	[illegible]	351K
Feature, Line Review		46K
Entry/Edit		14K
AIDs		45K
[illegible], BASELINE PLUS [illegible] INPUT:		
Session Control/Mode Set	765K	765K
Feature, Line Review		41K
Entry/Edit		33K
AIDs		43K
ON-LINE, ANALYST, BASELINE PLUS VOICE INPUT PLUS VOICE RESPONSE:		
Session Control/Mode Set	765K	825K
Feature, Line Review		41K
Entry/Edit		90K
AIDs		43K

Cost

Table [illegible] hardware costs for the two-level centralized con[illegible] amounts to roughly $9.1K per analyst. The principal [illegible] the reduction in hardware cost from the centralized configuration [illegible] Section 7.1 is the lower cost of the speech recognition equipment. As mentioned previously, non-hardware costs such as system engineering and software development are not included. However, these costs can be expected to be higher than those for the centralized configuration because of the greater complexity of this configuration.

7.3 INDEPENDENT STATION CONFIGURATION

Figure 19 illustrates a final system configuration consisting of completely independent stations. This approach represents the ultimate distribution of processing capability, and is purposely intended to be more austere than the preceding configurations. However, it boasts the greatest system redundancy, reliability, and flexibility.

The computer subsystem is implemented by means of a microcomputer having dual floppy disk drives as a means of on-line permanent storage rather than the rigid disks of the preceding configurations. The lower reliability of floppy disk drives relative to fixed disks is outweighed by the overall level of system reliability and availability, as a single point failure in any subsystem can do no more than disable a single analyst's station. Two double-density disk drives can be used to accommodate the on-line storage requirements summarized in Table 20. A total of 450K bytes of disk storage and 100K bytes of main memory are needed to allow at least 25 percent reserve storage capacity.

The analyst uses the system in a private manner. He begins the log-on process by inserting a disk containing his/her FADT files and voice reference patterns into one of the drives. The other drive contains the operating system, application software, and voice response samples. When he/she is ready to submit an FADT file or needs a printout, the analyst must remove the disk and transfer it manually to a separate station containing a tape drive and printer. Two such stations are required in this approach. Since it is likely that they will be heavily occupied, it is assumed that they would only be marginally useful for data entry.

TABLE 19. TYPICAL HARDWARE COSTS, TWO-LEVEL CENTRALIZED CONFIGURATION

Item	Basis	Cost
COMPUTER SUBSYSTEM	$205K	$205K
ANALYST STATIONS	$1K x 100	$100K
OPERATOR STATION	$1K x 1	$ 1K
VOICE RECOGNITION SUBSYSTEM	$3K x 100	$300K
VOICE RESPONSE SUBSYSTEM	$3K x 100	$300K
TOTAL		$906K

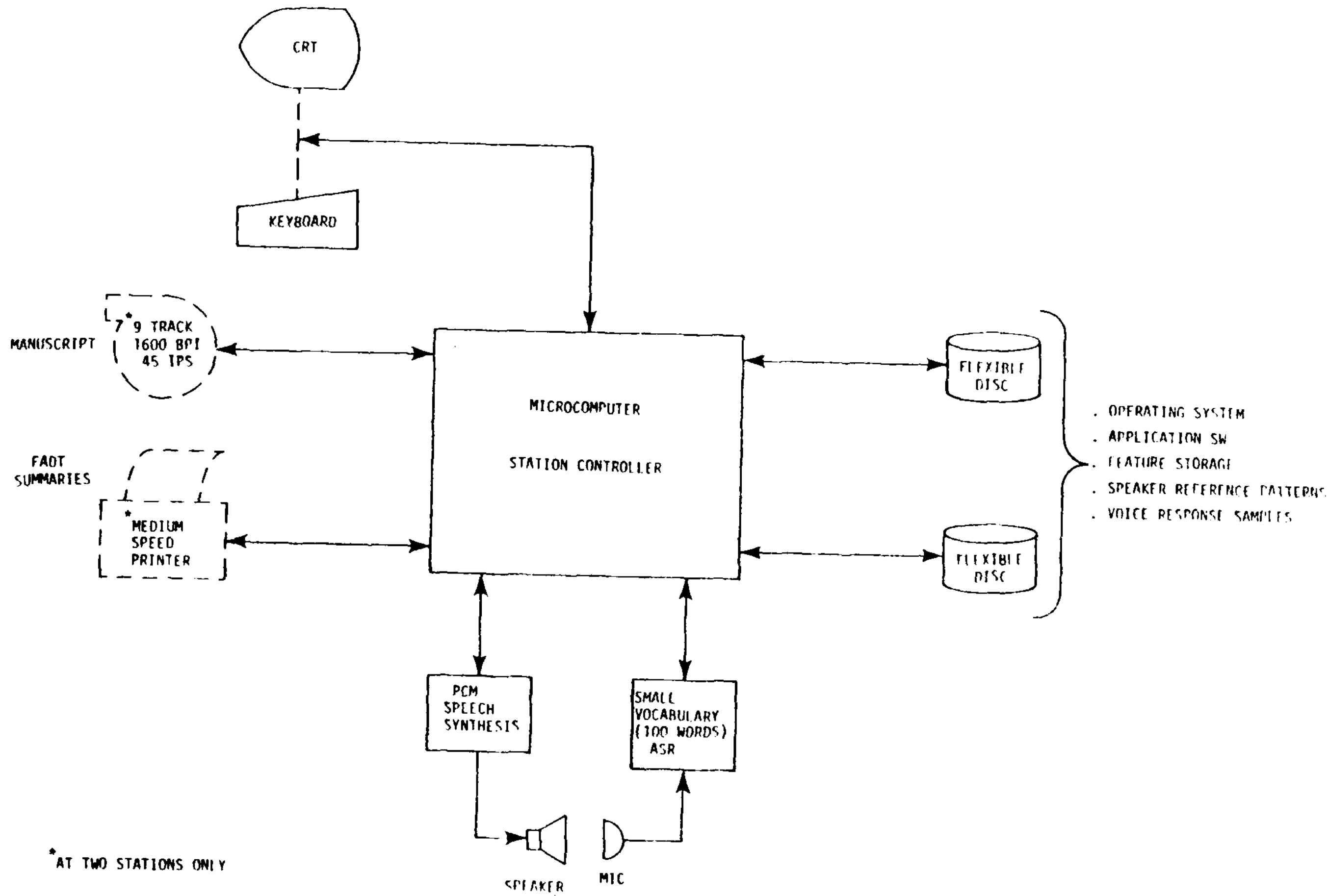


Figure 19. DLMS voice data entry system: independent stations approach.

TABLE 20. ESTIMATED ON-LINE STORAGE REQUIREMENTS, INDEPENDENT STATIONS CONFIGURATION

Category	Basis	Main Memory	Disk Memory
OPERATING SYSTEM	Source + Object + Load Module		200K bytes
	Load Module + Buffers	40K bytes	
APPLICATION SOFTWARE	Source + Object + Load Module		100K bytes
	Load Module + Buffers	40K bytes	
SPEAKER REFERENCE PATTERNS	300 words x 128 bytes		38.5K bytes
FEATURE STORAGE	2 files x 100 features x 30 characters		6K bytes
TOTALS		80K bytes	345K bytes

The voice recognition subsystem consists of a small-vocabulary speech recognition device such as that used in the configuration of Subsection 7.2 and its operation is essentially the same. A 1920 character per second serial interface services this device.

We shall not attempt to make specific performance estimates for this system, since it is not a multiple-user environment where competition for resources exists. A general-purpose microcomputer suitable for this design is capable of executing on the order of 250K instructions per second, giving this configuration a performance approaching that of the minicomputer-based controller configurations shown in the previous subsections.

Typical hardware costs for this system are estimated in Table 21. The average hardware cost per analyst station is about $12.4K (assuming 100 analyst stations).

7.4 VOICE RESPONSE SUBSYSTEMS

No mention was made of the specific type of voice response subsystem used in the three different configurations, and generally, all three types of speech synthesis methods mentioned in Section 3.2 can be used. The hardware costs of each type are essentially equivalent (as shown in Tables 14, 17, and 20); however, their quality and storage requirements are not. Voice output quality was discussed in Subsections 3.2.1, 3.2.2, and 3.2.3, and storage requirements are presented in Table 22.

For phoneme synthesis, the voice response subsystem would consist of a number of independent voice response terminals. Each terminal has a repertoire of phonemes (the primitive sounds contained in speech), and the computer subsystem causes the terminal to utter a phoneme by sending it a command, two characters in length. Words are uttered by concatenating phonemes. Thus, storage of the response vocabulary is actually split between the voice response terminal, which stores sounds, and the computer subsystem, which stores all the required words and phrases in terms of phonemes.

In general, there are as many phonemes in a word as there are letters, which corresponds to a maximum of 36 for the longest utterances in the spoken vocabulary. In normal speaking voice, it takes about 2 seconds to make the longest utterance, so the maximum data rate required is approximately 36 characters per second, assuming 2 characters per phoneme (as required for

TABLE 21. TYPICAL HARDWARE COSTS: INDEPENDENT STATIONS CONFIGURATION

Item	Basic Station Costs (100 Stations	Tape/Printer Station Costs (2 Stations)
Keyboard/CRT Terminal incl. Computer Subsystem and Floppy Disc Drives	$600K	$44K
Voice Recognition Equipment	$300K	
Voice Response Equipment	$300K	
TOTALS	$1200K	$44K

TABLE 22. STORAGE REQUIREMENTS FOR DLMS VOICE RESPONSE

Type of Speech Synthesis	Disk Storage Requirements
Simple Encoding (PCM)	1200K bytes
Complex Encoding (LPC)	360K bytes
Phoneme Synthesis	10.8K bytes

VOTRAX equipment). Thus, operation of the interface at a standard rate of 60 characters per second would meet this requirement.

The phoneme synthesis technique has the greatest advantage in the centralized configuration because of the expected high loading of its single minicomputer subsystem. It has the drawback of producing unnatural speech quality, however.

Since a voice response subsystem utilizing linear predictive coding (LPC) for speech synthesis allows for a much more natural speech quality than does phoneme synthesis, it is preferred when data rates are not critical. This tradeoff occurs as a result of the increase in data rate at the interface from approximately 40 characters per second for phoneme synthesis to about 1200 characters per second for LPC synthesis.

PCM speech synthesis affords the highest quality of speech. However, the high data rate involved, approximately 4K bytes per second, requires a parallel interface from the computer storage of voice output data to the PCM decoder. Thus high-quality PCM speech synthesis is actually best suited to the independent stations configuration, because of the relative simplicity of obtaining a parallel interface to the terminals microcomputer. The large storage requirement for PCM speech samples is met by the floppy disk. Since access time for the disk is under 250 msec, and the data transfer rate approaches 30K bytes per second, a 1-second utterance can be retrieved and playback begun in under 1 second.

7.5 CONFIGURATION EVALUATION

The quantitative performance data presented in this section confirm that all three configurations are viable. Unfortunately, no such quantitative data exists for the evaluation criteria presented in Section 4. Therefore, the latter criteria can only be evaluated qualitatively among the three configurations.

Several important criteria, such as data entry speed and accuracy, vocabulary flexibility, training requirements, and user acceptance in general, are largely a function of specific hardware and software capabilities and their implementation. For example, the type of headset used will play a strong role in affecting user acceptance. In some ways it is the most visible aspect of the VDE system from the analyst's viewpoint.

Such hardware and software capabilities are not configuration specific, but other criteria, such as the system's reliability and flexibility are. The independent stations configuration is clearly the most flexible and reliable--stations can be added and used at will, and even the total breakdown of a station would only affect a single analyst--but it is difficult to judge just how flexible and reliable the other configurations would be by comparison.

Our analysis showed that the computer subsystems in the two centralized configurations are only lightly loaded even under peak usage conditions. Hence, the only restrictions to adding stations are trivial hardware concerns, such as the maximum number of output ports available on the central controller(s). In this respect, the flexibility of the two-level configuration is greater than that of the centralized configuration, but either is adequate for the DLMS application.

The greatest weakness of the centralized configuration is its susceptibility to single-point failures in the central minicomputer. Depending on such factors as the average temperature and humidity, temperature variations, and cleanliness of the computer's environment, it will be more or less reliable. Nonetheless, the problem is that any downtime at all for the central computer will halt all analyst data entry.

Of course, data could always just be written down on paper at such times for entry later on when the computer is up again. However, this solution may not be acceptable in practice for an on-line, production system. Alternatively, additional memory could be included at each analyst's station for just such an eventuality. This memory could provide a few hours of temporary off-line data storage to allow the analysts to continue recording data in the interim.

The physical size requirements at the analyst's station are essentially the same for all three configurations: a keyboard/CRT terminal, and possibly an extra box containing voice recognition and response hardware, if these cannot be accommodated in extra card slots within the terminal. Terminals with additional card slots, such as microcomputer development systems, are generally larger than typical CRT terminals, but would have advantages in terms of reliability, development time and risk, and safety.

In terms of hardware costs, all three mixed-mode VDE configurations are marginally cost-effective, and to these figures must be added substantial costs for hardware and software design, development, and integration, not to mention implementation and maintenance costs. The baseline system alone (i.e., just keyboard/CRT terminal data entry) is 20 to 50 percent lower in hardware cost than any of the mixed-mode configurations.

Without significant justification for the advantages of voice data entry in this application, it is difficult to justify the additional costs it would incur. Thus, the independent stations configuration presents the greatest promise for mixed-mode VDE, but its costs are likely to be considered prohibitive.

8. CONCLUSIONS AND RECOMMENDATIONS

This configuration study presented both unique problems and unique opportunities in examining a potential application of voice data entry (VDE). The usefulness of data entry and verification by voice would seem to be obvious in the labor-intensive, manual, off-line processes now used for data entry to the DLMS data base. But it is difficult to quantitatively evaluate VDE under these circumstances. Nor is it clear how such subjective factors as analyst concentration and fatigue can be measured adequately. Completing OPSCAN forms and keypunching are poor examples of man-machine interaction and wasteful of the analyst's unique skills, but it is difficult to appreciate the value of on-line system capabilities from a strictly off-line viewpoint.

That the present system of data entry is outmoded and in need of improvement is evidenced by the Interactive Feature Analysis Support System (IFASS) and the Computer Assisted Photo Interpretation (CAPI) system due to be implemented at DMAAC. In light of these planned procurements, our conclusion is that the use of voice data entry is not an all-or-nothing choice and that full-scale implementation of VDE is unwarranted at DMAAC at this time. For this same reason, there is no point in committing to specific hardware for the configurations presented in Section 7.

From the experience of DMAAC with the advanced development mode [illegible] system, it would appear that keyboard data entry is superior [illegible] over it. But the evaluation tests consisted of [illegible] sample size, and we conclude that further tests [illegible] need to be performed. We further conclude [illegible] testing can best be accomplished using the [illegible] capabilities of IFASS and CAPI.

In this [illegible] to attach stand-alone voice input and [illegible] IFASS terminals to allow for mixed-mode [illegible] system will be invisible to the central IFASS [illegible] keyboard data entries will appear to be identical from [illegible] viewpoint. In fact, more than one VDE manufacturer is involved [illegible] their own mixed-mode data entry stations, and announcements on [illegible] can be expected this year.

The technological leap to VDE from off-line, manual data entry has been successfully made in a number of production environments, but dependence on

VDE is neither appropriate nor cost-effective for this application at this time. Currently, data entry comprises only a small portion of the analyst's time. However, reductions in the analysis workload resulting from the installation of CAPI workstations could, in comparison, easily double the percentage of the analyst's time spent on data entry. Nonetheless, VDE equipment does pose practical problems in the DMAAC environment, as discussed in previou[illegible] Since on-line keyboard data entry offers significant improveme[illegible] low costs, it presents the best alternative at this sta[illegible]nded.

Any advantages of VDE over keyboard data e[illegible]oncentration, hands-free/eyes-free data entry capab[illegible]eed and accuracy, and data coding requirements [illegible]parent only after on-line data-entry and [illegible] have been established, and after further ex[illegible] VDE capabilities. We believe that both th[illegible] a mixed-mode data entry system that will allow [illegible] as limited-capability VDE. Although the advantages [illegible] implementation will not be available, the cost and training [illegible] of mixed-mode VDE will be low enough to allow for implementation, [illegible], and experience on a larger, more practical scale. Without such experience, the practicality of VDE for DLMS data base entry is unclear and subject to conjecture.

With this in mind, our recommendations for future work are as follows:

First, procedures for on-line, real-time data entry and verification for the DLMS data base should be developed. These procedures should include the specification of probable data entry rates and realistic data response times. The development of such on-line data entry procedures will require a thorough examination of analysis tasks to arrive at the best analyst-computer interaction.

Second, in conjunction with the IFASS keyboard/keypad equipment, design a mixed-mode data entry station using state-of-the-art speech recognition technology. This station will be able to accept voice and keyboard data interchangeably, at any time, according to analyst desires. Advances in VDE equipment are occurring at such a rapid pace that the survey of devices presented in Section 3 will unavoidably become somewhat dated by the time this report is distributed.

Third, in light of the present coding procedures for DLMS features, implement and test limited-capability (20-to-30-word vocabulary) VDE for data entry speed and data accuracy in DFAD compilation under realistic analysis conditions. Feature ID and other codes will be entered as numeric codes rather than as verbal descriptions, thus improving recognition accuracy while greatly reducing analyst training and retraining times. The reduced training requirements should allow for more testing using more analysts entering more data under more realistic analysis conditions than were provided for in the ADM evaluation tests.

Fourth, in light of the nuisance and fatigue involved in wearing even lightweight headsets, evaluate comfortable, nonobstructing earpiece microphones as replacements for the close-talking headset-mounted microphones typically required for VDE equipment. If they could perform acceptably,* such hearing-aid-type microphones could provide a considerable advantage for VDE in Defense Mapping Agency applications which require stereoscope viewing. Certainly, their vastly improved comfort over headsets should add significantly to user acceptance of VDE.

Finally, a thorough examination of high-speed, high-quality voice response technology for prompt, feedback, confirmation, and verification functions associated with VDE should be performed. If designed and implemented properly, voice response should improve data entry speed while reducing the number of errors. The speech synthesis device used in the ADM evaluation tests represents only one type of voice response equipment available, and it is questionable whether it was, in fact, working properly. Hence, the value of voice response capability should not be judged on this experience alone.

---

*Several VDE manufacturers have informally tested Lear Siegler's EarCom device, but their experience is not applicable for several reasons, including the fact that they did not make use of the custom-fitted earpieces, without which the device provides essentially no attenuation of extraneous noises. In addition, other devices and wider bandwidth transducers should also be evaluated.

REFERENCES

Aviation Week & Space Technology, "Word Recognition System Cuts Work in Parts Training," 15 September 1980, pp. 74-75, 79.

Defense Mapping Agency, Product Specification for Digital Landmass System (DLMS) Data Base, Defense Mapping Agency Aerospace Center, St. Louis AFS, Missouri, July 1977.

Kleinrock, L., Queuing Systems, Volume II: Computer Applications, John Wiley & Sons, New York, 1976.

Scott, Phillips B., Final Technical Report: DLMS Voice Data Entry, Contract F30602-78-C-0327, Rome Air Development Center, Griffiss AFB, New York, 1980.

Welch, J. R., Automatic Data Entry Analysis, Final Technical Report RADC-TR-77-306, Rome Air Development Center, Griffiss AFB, New York, 1977.

Appendix A

FADT SUMMARY

The DLMS Culture File (DFAD) is created directly from feature manuscripts. Manuscripts of various geographical areas are the building blocks of DFAD. The information in DFAD is that recorded for each manuscript in a Feature Analysis Data Table (FADT) (Figure A-1).

A	B	C	D	•E	•F	•G	H	I	• J_1	J_2	K	L	M
FAC NO.	FEATURE TYPE	SURFACE MATERIAL	PREDOMINANT HEIGHT	STRUCTURES PER SQ. N.M.	PERCENT OF TREES	PERCENT OF ROOF	FEATURE IDENT.	DIRECTIVITY/ ORIENTATION	LENGTH OR DIAMETER OF POINT FEATURES	WIDTH OF LINE OR POINT FEATURES	LEVEL	NUMBER OF PYLONS	SHEET IDENTIFICATION
1 2 3 4	6	8 9	11 12 13 14	16 17	19 20 21	23 24 25	27 28 29	31 32 33	35 36 37	39 40 41	43	45 46 47	49 50 51

Figure A-1. Feature Analysis Data Table.

Each column represents a digital code recording a parameter for that row's respective feature. Not all columns are filled in for each row. One indicant of which are required is Feature Type, which is recorded in column B. Feature type may be one of three codes:

0 - A point feature, represented by a dot on a map;

1 - A line feature, represented by a line on a map;

2 - An areal feature, represented as a closed figure on a map.

Depending on feature type, other column values are recorded according to the table in Figure A-2. A discussion of the attributes in each of the columns follows.

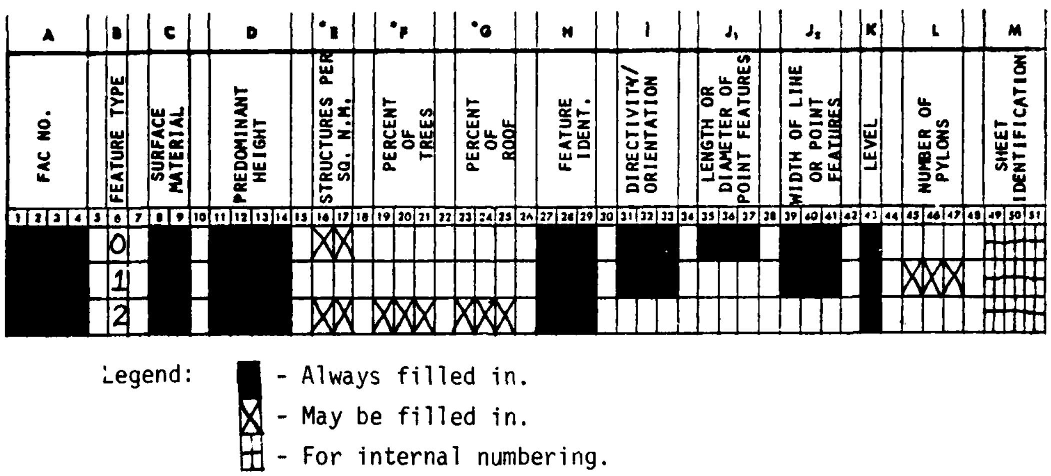


Figure A-2. Sample recording of column values.

Column A contains the Feature Analysis Code(FAC) number. This is a unique number assigned for each feature in the manuscript. The numbers begin with 1 and continue sequentially up to a possible 9999. The average manuscript contains about 500 features. These numbers serve as an index for the features they describe.

Column C contains the feature's Surface Material Code (SMC). This indicates what the feature is generally made of. Codes are integer values from 1 to 13, representing the following materials:

SMC	Material
1	Metal
2	Part Metal
3	Stone/Brick
4	Composition
5	Earthen Works
6	Water
7	Desert/Sand
8	Rock
9	Asphalt/Concrete
10	Soil
11	Marsh
12	Trees
13	Snow/Ice

Column D contains the feature's Predominant Height. The feature's height above ground level is estimated and mapped to the nearest 2-meter increment between -1022 and 1022 meters. Maximum code length is four digits.

Columns E, F, and G are only filled in for areal features with Surface Material Codes less than 4, and in one special case, for point features. Column E contains Structures per Square Nautical Mile, an estimate of structure density; Column G contains the Percent of Roof Cover of the area, based on E. Both of these are filled in according to Table A-1. Column F contains an estimate of tree density, Percent of Tree Coverage. It contains one of three values (0, 10, 30), each a percentage.

TABLE A-1. NUMBER OF STRUCTURES AND PERCENT OF ROOF COVER

Number of Structures Per Square Kilometer	Number of Structures Per Square Nautical Mile	Number of Structures Code Number	Percent of Roof Cover
1	1	0	100
2 - 105	2 - 365	1	30
106 - 252	366 - 865	2	30
253 - 544	866 - 1865	3	30
545 - 679	1866 - 2330	4	30
680 - 825	2331 - 2830	5	20
826 - 971	2831 - 3330	6	20
972 - 1114	3331 - 3830	7	20
1115 - 1262	3831 - 4330	8	20
1263 - 1408	4331 - 4830	9	20
1409 - 1554	4831 - 5330	10	20
1555 - 1700	5331 - 5830	11	20
1701 - 1845	5831 - 6330	12	20
More than 1845	More than 6330	13	20

Column H contains the Feature Identification Code (FIC), the three-digit code which indicates just what the feature is. The DMA Product Specification for DLMS Data Base lists 255 features and their codes. They are organized by first digit into the following groups:

100	Industry
200	Transportation
300	Commercial/Recreation
400	Residential/Agricultural
500	Communications/Transmission Facilities
600	Government and Institutional
700	Military/Civil Installations
800	Storage
900	Landforms, Vegetation, and Miscellaneous

Column I represents one of two things:

1. For a point feature, its Orientation, the angle from true north to its major axis. Values are either rounded to the nearest 5° or entered as codes from 0 to 31, where the code multiplied by 11.25° is the nearest angle, or as 63, which represents omni-directivity.

2. For a line feature, it represents Directivity of the side with greatest radar reflectivity. The codes are three digits, with

   001 - Uni-directional
   002 - Bi-directional
   003 - Omni-directional

Columns J_1 and J_2 contain dimensions of features. Each holds three digits, representing distances rounded to 2-meter increments between 0 and 254 meters. J_1 is used for the length of point features; J_2 is used for the width of either point or line features.

Column K contains the level of the feature analysis being performed for this manuscript. There are two levels:

Analysis Level	Code No.
Level 1	1
Level 2	2

where Level 2 requires finer analysis to be performed. Since manuscript level is constant, an ASR system would only require this value once, at the beginning of analysis.

Column L holds Number of Pylons, an infrequently used three-digit value that does not appear in the DFAD product.

Column M holds Sheet Identification numbers. These comprise an internal numbering system that does not appear in the final file. An ASR system could similarly require only one value at set-up time to eliminate these values.

Appendix B

QUESTIONNAIRE RESULTS

In late April 1980, TSC created a six-page questionnaire, to be completed by each analyst at DMAAC, and submitted it to DMAAC. The questionnaire was part of our task to analyze and evaluate the present operational procedures for the automated compilation of the Feature Analysis Data Table (FADT) for the Digital Landmass System (DLMS) and our task to analyze the feasibility of, and make recommendations for, integrating voice recognition technology into the existing DMAAC system.

The questionnaire asked 14 multiple-choice questions, some with multiple parts, and provided space for general comments. Eighty-four analysts completed the questionnaire. A tabulation of their responses to each question is shown in the copy of the questionnaire at the end of this appendix.

The first two questions deal with the analyst's length of experience performing feature analysis at DMAAC. It turned out that roughly three-fourths of the respondents have been performing DFAD analysis for two or more years, and one-half for more than four years.

The next three questions sought to determine how familiar each analyst was with surface material, areal feature, and feature ID codes. Since there are about 260 feature ID codes (FICs) but only 13 surface material codes (SMCs), we expected the analysts to be less familiar with the former than the latter, which was generally the case. Over three-fourths of the analysts were very familiar with SMCs and only rarely needed to refer to documentation, while little more than one-half of them felt this way about FICs. About 10 percent of the analysts thought they had SMCs and areal feature codes memorized, but no one admitted to having the FICs memorized; in general, a significant portion of the analysts said they had to refer to documentation for FICs fairly often.

After giving the analysts a few questions (questions 6, 7, and 8) about their manual methods of data entry, how they would evaluate it, and the types of errors that occur, we questioned them about how their data entry procedures might be affected by voice data entry, questions 9, 10, 11, and part of 12. In those questions we introduced the concept of a personal secretary who would provide "flawless" data recording for each analyst. We considered this the best way to present a "perfect" voice data entry system for evaluation in realistic terms without getting bogged down in its technical aspects.

Close to 75 percent of the analysts responded that the secretary concept, that is, a perfect voice data entry system, would make analysis easier (question 9A) and their work faster (question 9B). Fully two-thirds felt that they could work more efficiently and with fewer errors (question 9C). Most of the remaining analysts answered that their work would not be significantly affected one way or another.

One-half the analysts indicated that they would prefer the freedom to either say a feature's name or specify its appropriate FIC (question 10). The remaining analysts were split evenly into those who would always prefer to specify the FIC and those who would always prefer to say the feature's name. On the other hand, almost two-thirds of the analysts said that they would prefer to always specify SMCs rather than describe the surface material itself (question 11). No doubt this reflects most analysts' familiarity with the 13 surface material codes (SMCs). In fact, only about 10 percent of the analysts indicated that they would always rather say the surface material name than its code.

Questions 13 and 14 sought to determine the analysts' familiarity with VDE equipment as well as with other computer peripheral devices, and judging from this experience, their opinion of the practicality of such equipment.

No more than one-third of the analysts had ever seen a speech recognition device demonstrated (question 13A), and only one-half had any experience with CRT terminals (14A). Most of the analysts who had seen speech recognition devices demonstrated were referring to the ADM VDE system at DMAAC, and only three of those who responded had any experience with it themselves.

Despite the general lack of experience with speech recognition devices and other on-line data entry devices, well over one-half of the analysts were favorably inclined toward voice data entry. That is, most analysts considered that it would be an improvement over the current system. Given the analysts' relative lack of experience with VDE equipment, this opinion would seem to reflect some measure of their dissatisfaction with the current system.

One-fourth of the analysts answered that data entry by voice was either impractical or would degrade performance over the present system (question 13C). Some of the analysts' comments are illuminating in this regard; the more perceptive and feisty comments have been included in Table B-1. As to the responses to question 13D, one-half of the analysts stated that they would have no objections to having to wear a headset, provided that it was lightweight and comfortable. However, almost one-fourth of them were not sure, and an additional one-fourth did not want to have to wear headsets regardless oi whether they were comfortable.

In addition to straightforwardly tabulating the data, we also wanted to analyze the correlations between various questions. For example, Were the analysts who felt that voice data entry would be an improvement those analysts who had been working on DFAD analysis for many years or only a relatively short time? We found that experience did not seem to make any difference; the same range of opinions about the practicality and usefulness of VDE was held regardless of the analyst's years of DFAD analysis.

However, the correlation between some questions showed a definite trend. Not surprisingly, analysts who answered that current methods of data entry were pretty good did not feel that voice data entry would be much of an improvement, and vice versa. Similarly, analysts who considered an ideal VDE system (the secretary concept) to be advantageous also answered that data entry by voice would be a practical improvement over the present system of data entry. Only about 10 percent of the analysts indicated that, although the concept of voice data entry was a good one, in practice it would turn out to be worse than present methods.

Of the analysts who had seen a VDE device demonstrated, two-thirds thought that voice data entry would be better than current methods of data entry. Of those who had never seen a VDE device demonstrated, opinion was fairly evenly split between those who thought that VDE would be an improvement and those who checked that it would make no difference or in fact be worse than the current methods of data entry. In general, only one out of every three analysts had seen voice data entry demonstrated.

Just over one-half of the analysts had experience with CRT terminals, and of these, three-fourths indicated that voice data entry would be an improvement over the present system. Of those who had no experience with CRTs, opinion was split evenly between those who checked that VDE would be better and those who checked that it would be worse.

Lastly, of the analysts who thought that voice data entry would be an improvement over the present system of data entry, four out of every five had no objections to wearing a headset, at least provided that it was lightweight and comfortable. The 25 percent of the analyst population who indicated that VDF would be worse than current methods did not seem to care one way or another about wearing headsets.

TABLE B-1. SAMPLE COMMENTS FROM THE QUESTIONNAIRES

"The value of such a speech recognition system will depend on how it is integrated into the production system."

"Eliminating the necessity for manually recording seems to be more efficient, but undesirably restrictive on the operator."

"I think it's a good theory; however, the expense of the hardware and user training may render it impractical. Cost efficiency may make a little keypunching or opscan rework relatively minor by comparison."

"When electronic equipment is inoperative, I am out of work."

"I think this is just another case of automation for the sake of automation."

"I find it hard to evaluate a piece of equipment and its benefits without ever having seen or used it. It seems like there might be a more efficient way to record data without all the complexity being suggested. Maintenance on the equipment we have is very poor without creating a larger work load of equipment. The purpose of innovation is to simplify not to create greater chaos."

"The only advantage to buying this machine would be gained by a stockholder in the company."

"I doubt the efficiency obtained with this system would make a significant improvement in production, much less be cost-effective. For our purposes speed is more easily maintained when you can have a graphic record of outlines, FAC numbers, and data in front of your eyes. Having to call up information when you lose your place or your concentration would be a hindrance to me. Manual recording of data along with some type of visual record would be preferable to me, such as a keyboard with a CRT."

"I think voice recognition systems are inevitable and will be a technological boon once an adequate level of reliability has been reached. Such technology isn't cheap, however, and the early work in this area will not produce immediate dividends. Perhaps the government does have an economic responsibility to support such work through its early stages; but I do not believe the argument is valid at the production level in CDI. I think voice systems would speed up our work, but not enough to justify the greatly increased expense. I would retain one or two stations as a testing device for the next several years until refinements have reached a stage which allows cartographers to use the system without requiring individual voice imprints."

TSC QUESTIONNAIRE: TABULATION OF ANALYST RESPONSES (84 RESPONDENTS)

1. How long have you been performing scene analysis and/or feature extraction at DMAAC?

   [1] under 6 months
   [23] 6 months to 2 years
   [29] 2 to 4 years
   [31] more than 4 years

2. How long have you been performing DFAD cultural analysis?

   [2] under 6 months
   [22] 6 months to 2 years
   [30] 2 to 4 years
   [30] more than 4 years

3. Indicate your familiarization with DLMS feature identification codes (FIC's) in your work.

   1 [0] codes memorized, never need to refer to documentation
   2 [45] very familiar with codes, only rarely need to refer to documentation
   3 [38] moderately familiar with codes, need to refer to documentation fairly frequently
   4 [1] not very familiar, need to refer to documentation very often
   5 [0] unfamiliar, always need to refer to documentation

4. Using same scale as in Question 3, indicate your familiarization with surface material codes.

   1 [12] codes memorized, never need to refer to documentation
   2 [63] very familiar with codes, only rarely need to refer to documentation
   3 [9] moderately familiar with codes, need to refer to documentation fairly frequently
   4 [0] not very familiar, need to refer to documentation very often
   5 [0] unfamiliar, always need to refer to documentation

5. Using the same scale as in Question 3, indicate your familiarization with the areal feature codes; that is, structures per square mile, percent of trees, and percent of roof.

   1 [7] codes memorized, never need to refer to documentation

   2 [38] very familiar with codes, only rarely need to refer to documentation

   3 [28] moderately familiar with codes, need to refer to documentation fairly frequently

   4 [11] not very familiar, need to refer to documentation very often

   5 [0] unfamiliar, always need to refer to documentation

6. Indicate your current method of recording feature data manually.

   [57] FADT keypunch forms   [41] OP-SCAN FADT (Form 5600/AC1-20) forms

   [11] other (describe) ______________________________

7. Questions A-D list some qualities which could describe your current data entry method in relation to your other work activities. Please indicate your opinion by checking the appropriate box.

   A. Time efficiency. The current method of data entry...

   [32] Helps same time   [43] Consumes too much time   [8] No opinion

   B. Concentration. The current method...

   [46] Fits in smoothly with my work   [28] Makes me lose my concentration   [8] No opinion

   C. Physical requirements. The current method...

   [58] Poses no movement problems   [19] Requires me to move about my work station excessively   [7] No opinion

   D. Adjustment. The current method...

   [68] Is easy to learn and perform   [9] Is confusing and/or complex to learn and perform   [5] No opinion

8. With regards to error making, indicate if the errors listed below occur in the current method of recording feature data.

incorrect numeric code written down	[3] frequently	[73] occasionally	[6] never
code scanned/read incorrectly	[5] frequently	[60] occasionally	[12] never
code cannot be scanned/read	[0] frequently	[49] occasionally	[23] never

Please indicate other errors that occur:

____

____

For Questions 9 through 12, imagine that you had your own personal secretary, who would write down all the feature data, without errors, onto FADT forms as you said it to him/her; that is, the secretary would free you from having to record data manually.

9. Would having this secretary make your work...

A. EASIER?

$\boxed{33}$ 1 A lot easier	$\boxed{28}$ 2 Somewhat easier	$\boxed{18}$ 3 Not any easier	$\boxed{2}$ 4 Somewhat more difficult	$\boxed{2}$ 5 A lot more difficult

Why? ______________________________

B. GO FASTER?

$\boxed{32}$ 1 A lot faster	$\boxed{27}$ 2 Somewhat faster	$\boxed{15}$ 3 Not any faster	$\boxed{5}$ 4 Somewhat slower	$\boxed{1}$ 5 A lot slower

Why? ______________________________

C. MORE EFFICIENT?

$\boxed{25}$ 1 A lot more efficient	$\boxed{28}$ 2 Somewhat more efficient	$\boxed{21}$ 3 Not any more efficient	$\boxed{6}$ 4 Somewhat less efficient	$\boxed{0}$ 5 A lot less efficient

Why? ______________________________

D. LESS DISTRACTING?

$\boxed{19}$ 1 It would be a lot easier to concentrate	$\boxed{21}$ 2 It would be somewhat easier to concentrate	$\boxed{25}$ 3 My concentration would not be affected
$\boxed{13}$ 4 It would be somewhat more difficult to concentrate	$\boxed{2}$ 5 It would be be a lot more difficult to concentrate	

Why? ______________________________

E. CONTAIN FEWER ERRORS?

$\boxed{19}$ 1 A lot fewer errors	$\boxed{34}$ 2 Somewhat fewer errors	$\boxed{21}$ 3 No fewer errors	$\boxed{6}$ 4 Somewhat more errors	$\boxed{0}$ 5 A lot more errors

Why? ______________________________

10. In telling feature identification information to the secretary, which of the following choices would be most natural for you?

    [21] always saying the feature identification code, for example: "944"

    [21] always saying the feature's identification definition, for example "waterfall", and letting the secretary translate it into the code.

    [39] sometimes saying the feature's identification definition and sometimes saying the code

11. In telling the secretary surface material information, which would be most natural?

    [52] always saying the surface material code

    [11] always saying the name of the surface material category

    [19] sometimes saying the name of the surface material category and sometimes saying the surface material code

12. *This question deals with the amount of time* it takes for you to derive all of the information about one feature (i.e., feature type, feature identification, predominant height, direction, length, etc.)

    A. In reporting to the secretary all of the information about one feature, would it be more natural for you to report (choose one)

    [29] all of the necessary data at one point in time

    [50] individual data items as you determine them, perhaps requiring several minutes to report all of the information about the feature

    [2] no opinion

    B. Ignoring the secretary concept, which of these choices in 12.A describes how you currently record data manually?

    [29] all the data at once

    [48] individual data items as you determine them

    [8] neither of these (please explain): ____________________

    ____

    ____

    ____

13. At present, a speech recognition machine is being evaluated for application to enter FADT information by voice.

A. Have you ever used a speech recognition device?

[3] Yes, the one being tested at DMAAC

[0] Yes, the one at DMAHTC

[0] Yes, another device (if possible, specify: ________________ ________________).

[79] No

B. Have you ever seen a speech recognition device demonstrated?

[22] Yes, here at DMAAC

[4] Yes, another device

[56] No

C. What is your guess as to the practicality of feature data entry by voice? (Choose one)

[15] 1 It would be a definite improvement over the current system	[32] 2 It would probably be better than the current system	[10] 3 It would make little difference	[14] 4 It would probably be worse than the current system	[6] 5 It would be a waste of time

Additional comments ________________

D. For such a system, would you have objections to wearing a headset?

[10] Not at all

[40] Not if its lightweight and comfortable

[14] Not sure

[18] Yes, please indicate why ________________

14. How much experience do you have operating computer peripheral devices?

A. CRT Terminals	B. Keypunch	C. Tape Drive	D. Other: ___________
34 None	1 None	47 None	25 None
22 A little	40 A little	15 A little	11 A little
15 A lot	42 A lot	10 A Lot	4 A lot

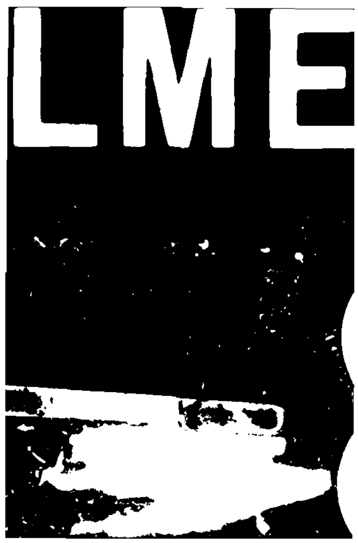
15. If you have any questions or remarks about the questions asked or answers provided in this questionnaire, or any additional comments about the issues raised herein, please write them in the remaining space. Use additional paper if necessary. If you are remarking about a specific question, please indicate the question number with your comments.

Name (OPTIONAL) ______________________________

THANK YOU FOR YOUR HELP.

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.


LME